The Mirror-Neuron Paradox: How Far is Sympathy from Compassion, Indulgence, and Adulation?

Khalil, Elias

Monash University

11 June 2007

The Mirror-Neuron Paradox:
How Far is Sympathy from Compassion, Indulgence, and Adulation?

Elias L. Khalil[1]

ABSTRACT

Mirror neurons become instigated when the spectator *empathizes* with the principal's intention. But when they involve *imitation*, empathy (understanding) is irrelevant. While understanding may attenuate the principal's emotion, imitation escalates it. Such contradictory attenuation/escalation pathways of fellow-feeling can be solved if we distinguish two axes. While "rationality axis" asks whether the action is efficient or suboptimal, "intentionality axis" asks whether the intention is wellbeing or evil. The solution shows how group solidarity differs from altruism and fairness; how revulsion differs from squeamishness; how sympathy differs from adulation; how evil differs from selfishness; and how racial hatred differs from racial segregation.

Keywords: Adam Smith; David Hume's Fellow-Feeling Paradox; Desire; Paris Hilton; Crankcase Oil Problem; Comprehension; Understanding (empathy or theory of mind); Imitation; Status Inequality; Elitism; Authority; Pity: Obsequiousness; Racial Segregation; Racial Hatred; Rationality Axis; Intentionality Axis; Propriety; Impropriety; Revulsion; Social Preferences; Altruism; *Assabiya* (group solidarity); Fairness; Schadenfreude (envy/spite/malevolence/evil); Vengeance

JEL Code: D01; D64

---

[1] Elias.khalil@buseco.monash.edu.au Department of Economics, Monash University, Clayton, Victoria, Australia. The paper was supported by the Konrad Lorenz Institute for Evolution and Cognition Research (Altenberg, Austria). During my stay at the Konrad Lorenz Institute, I benefited immensely from the very generous comments and extensive conversations with Riccardo Draghi-Lorenz. I also benefited greatly from communications with Julie Zilko. This version of the paper received comments from Aldo Rustichini, Ian McDonald, Caroline Gerschlager, Herbert Gintis, and Roland Cheo Kim San. A much earlier version received comments from Philippe Fontaine, Ulrich Krohs, Robert Sugden, seminar participants at the Konrad Lorenz Institute, George Mason University (Center for Public Choice), and Monash University. The usual caveat applies.

0. The Many Faces of Fellow-Feeling

In her only published novel, *To Kill a Mockingbird*, Harper Lee tells stories about everyday life and racial segregation in a backwoods town in the Deep American South. The novel takes place in 1932, Maycomb County, Alabama. Tom Robinson is an African-American young man wrongly accused and, without one iota of evidence, convicted of "raping" Mayella Violet Ewell, a 19-year old white woman. From his testimony in court, it was she who tried to seduce him and he actually fled her house. She knew him as he passed her house daily in his way to the fields. She repeatedly asked for help with various chores in the yard, which he happily complied without taking a penny from her. The prosecutor, Mr. Gilmer, leveled a barrage of questions as to why would Tom help the woman: "Why were you anxious to do that woman's chores"—with her father and seven children on the place? "You did all this chopping and work from sheer goodness, boy?"? "You're a mighty good fellow, it seems – did all this for not one penny?" [Lee, 1989, pp. 217].

Tom finally explained: "I felt right sorry for her, …." Sure enough, there are plenty of reasons to feel sorry for Mayella: her mother has long been dead, her father drank most of the relief check and abused her when drunk, and she was the oldest of so many younger siblings. But as soon as Tom uttered his words of fellow-feeling, he interrupted himself. He realized, even before Mr. Gilmer gleamed over his prize, that he made a big mistake:

> `*You* felt sorry for *her*, you felt *sorry* for her?' Mr. Gilmer seemed ready to rise to the ceiling.
>
> The witness realized his mistake and shifted uncomfortably in the chair. But the damage was done. … nobody liked Tom Robinson's answer. Mr. Gilmer paused a long time to let it sink in [Lee, 1989, pp. 218].

Tom definitely damaged his case. How could he have the galls, a black person, feel sorry for

a white person, as the emphasis in Mr. Gilmer's voice indicates?[2] If he felt any fellow-feeling, it should be *adulation* and even *obsequiousness* towards white people. And they, in return, would have *authority* and feel superior or even *pity* towards him. So, for Tom to claim that he felt sorrow for Mayella can only be interpreted by the white jury as pity, i.e., as what they feel towards black people. Even if Tom's fellow-feeling is empathy, empathy entails status equality. It is obvious to anybody, given the institutional matrix of status inequality and racial segregation, Tom's motive cannot be empathy—not to mention his story that a white woman tried to seduce a man of lower status. It should not be surprising, therefore, that the jury found Tom guilty of rape.

This paper uses the term "fellow-feeling" or "mirroring" as a primitive that is shared among the more complex emotions of "sympathy," "compassion," "indulgence," and "adulation." This paper promises to show the exact differences among these complex emotions. Of course, one can choose other terms for them. The task would still be the same: how could a primitive such as fellow-feeling give rise to adulation as opposed to sympathy under different institutional matrices. Actually, the primitive fellow-feeling has no meaning if abstracted from the institutional matrix such as social segregation. The institutional matrix allows fellow-feeling to take a multitude of recognizable faces. If one ignores the institutional matrix, one would be perplexed as to why Tom's fellow-feeling towards Mayella worked against him. It would be naïve to assume that the exchange of fellow-feeling among agents can be abstracted from status inequality, the context of the situation, and so on.[3]

---

[2] In the 1962 film version, by same title, the defense attorney, Mr. Gilmer inserted "a white woman": *"You* felt sorry for *her*, a white woman, you felt *sorry* for her?"

[3] Aside from exchange of fellow-feeling, what about the exchange of goods in a world of status

To see how context matters in the processing of fellow-feeling, let us examine a true story, the sudden death of David Rivas Morales, a 40-year old house painter, on 20th June 2007:

> **An angry Texas crowd has beaten and killed a 40-year-old car passenger after a driver injured a young girl near the site of a busy local festival.** Police said the driver of the car had stopped to check on the health of the girl, said to be aged three or four. But when the passenger got out to see how she was, he was set upon by a group of up to 20 people before being left lying in a car park, police said. The girl was hit at low speed and was not seriously injured.
> The incident happened near Austin, Texas, as crowd of between 2,000-3,000 people gathered for the annual Juneteenth festival, which commemorates the freeing of American slave ["US crowd beats passenger to death," www.BBC.com].

Mr. Morales was simply the passenger who got out of the automobile once the driver stopped to check on the child who was not seriously injured. According to Harold Piatt, from the Austin Police Department: "It's that same crowd mindset of being one face in 1,000. Things get out of hand pretty quickly and people don't have the good sense to stop" [*Ibid.*].

The case of Mr. Morales does not mean that we should always employ "good sense," i.e.,

---

inequality and racial segregation? Do they exchange according to cost of production irrespective of the status of the producers? Can the term of exchange be naïvely detached from the institutional matrix? Thorstein Veblen [1934] thought it would be naïve to assume that, e.g., a shirt is a shirt irrespective of who produced it. He argued that demand may increase if the price of a good, probably as the case with "brand names," rises because it would acquire a snob appeal. Neoclassical economic theorists started recently to recognize the snob appeal, but only to treat as a feature of the good. Classical economics in the form of labor theory of value, on the other hand, does not recognize snob appeal. For instance, Karl Marx's [1976, ch. 1] concept of "abstract labor" [see Khalil, 1992] is expressly advanced to deny the role of status inequality. The concept assumes that prices are determined by equal labor-time—irrespective if produced by high- or by low-ranking agents. Classical economics assumes naively that exchange of goods disregard the issue of status. That is, two goods that cost the same should sell for the same price in competitive markets—ignoring the role of brand names [see Ewing *et al.*, 2007]. This naïve assumption was challenged, although on unnecessarily repugnant racial and colonial elitism, by Thomas Carlyle in the 19th Century. As David Levy [2001] demonstrates, Carlyle dubbed economics the "dismal science" exactly because it ignores the role of status inequality in the

what is called below "understanding" in the sense of knowing the context or intentions of the original emotion. In other institutional context, "mob mentality" is actually required—as in the case of fashion where people derive pleasure from seeing others imitating their taste [Ewing *e al.,* 2007]. Also, in balls and parties, people consume alcohol exactly in order to loosen up or to suspend "good sense." Otherwise, if people are inhibited and do not replicate the pleasantness of each other, the party would be a meeting of analysis and discussions. But in Mr. Morales' case, the familiar imitation and escalation of the original fellow-feeling took place rather than attenuation. Even if he child was *ex ante* thought to have been seriously injured, the fellow-feeling could (and should) have taken the attenuation pathway.

Interestingly, David Hume [in Smith, 1977, p. 43] has long ago noticed that fellow-feeling can proceed along the escalation pathway, as seen in mob mentality, as opposed along the attenuation pathway, as when "good sense" prevails. Hume was puzzled by the coexistence of these two pathways and considered it a contradiction—what is called here the "fellow-feeling paradox." He challenged his friend, Adam Smith, with the paradox. Smith responded by developing a particular notion of sympathy that remarkably can account for the attenuation pathway. A few authors have noted how Smith's notion is remarkably similar to the recently discovered mirror neurons, the seat of fellow-feeling [see Rustichini, 2005; Rizzolatti & Craighero, 2004a].[4] This paper uses to a great extent Smith's notion of sympathy. It ultimately finds that Smith failed to solve Hume's fellow-feeling paradox. Namely, while Smith can account for the attenuation pathway of

---

exchange of products [see Khalil, 2007a].

[4]There is an intimate link between MNS (fellow-feeling) and interpersonal utility comparison [see Fontaine, 2001]. But this ramification of MNS is not pursued here.

fellow-feeling, he cannot account for the escalation pathway. That is, how could the two pathways coexist?

It is argued here that the fellow-feeling paradox, i.e., the coexistence of escalation and attenuation pathways, is intimately connected to a debate concerning the different functions of the mirror neuron system (MNS) and, generally, the canonical neuron system (CNS). These systems were discovered in particular regions of the brain of primates and other mammals. They are usually identified as the seats of fellow-feeling expressed by an observer (called "spectator") towards the action of an observed organism (called "principal"). The MNS of the spectator is instigated when the spectator *understands*, i.e., empathizes with the intention of the principal in reaction to incentive. The CNS of the spectator, but also MNS if the spectator is human, facilitate *imitation* whose function, by definition, ignores the intention of the principal, i.e., does not involve the function of understanding (empathy).

So, MNS and, more broadly CNS, are characterized by "imitation function," on one hand, and "understanding function," on the other. This paper shows that the "imitation function" gives rise the escalation pathway as noticed in mob behavior, parties, and fads. On the other hand, the "understanding function" originates the attenuation pathway as Smith's analysis indicates.

But still, this finding—the tracing of pathways to functions—amounts to re-labeling the fellow-feeling paradox: One still has to ask: What demarcates the "imitation function" from the "understanding function" of the neural system? The inability to locate the difference between imitation and understanding, and how one function of the neural system is engaged rather than the other, is called here he "mirror-neuron paradox." To trace Hume's fellow-feeling paradox to the

neural level simply postpones the question: Still, what exactly sets the same primitive, fellow-feeling, to act as imitation as opposed to understanding? What exactly sets "understanding" apart from "imitation"?

The central aim of this paper is to solve fellow-feeling paradox or, what is the same thing, the mirror-neuron paradox. The solution has many payoffs. To mention one, it offers a *single* theoretic frame that simultaneously explains Tom's conviction of rape (i.e., the exact difference between sympathy and pity) and Morales' death (i.e., the escalation as opposed to the attenuation pathways of fellow-feeling). The paper's argument proceeds along the following sections:

1. What is Hume's fellow-feeling paradox?
2. How is Hume's paradox related to the mirror-neuron paradox?
3. What is Adam Smith's solution of the paradox, and why it is unsuccessful?
4. What is the solution of the paradox?
5. Can we test the proposed solution, the two-axis of evaluation hypothesis (TAE)?
6. Can TAE show how to model the diverse faces of fellow-feeling such as "sympathy," "compassion," "indulgence," and "adulation"?
7. Can TAE shed light on tribalism and nationalism and distinguish group solidarity from altruism and fairness?

1. David Hume's Fellow-Feeling Paradox

Hume challenged his friend, Smith, with a paradox.[5] In his 28th July 1759 letter to Smith [1977, p. 43], Hume posed the following question: In most cases, the fellow-feeling with someone who is ill-humored or in a bad mood *escalates* the gloomy atmosphere. If so, how come fellow-feeling with a grieving parent usually *attenuates* sadness and engenders joy [see Khalil, 2007e]? It is not sufficient to state that we have two kinds of sympathies. This would only beg the question. Why would

[5] Eric Schliesser alerted me to the letter. David Levy and Sandra Peart [2004] and Robert Sugden [2002] discuss the letter.

fellow-feeling, for no apparent reason, *attenuate* the original emotion in one case and escalate it in another—regardless whether the original emotion is grief or joy? So, the paradox, which repeats itself with respect to the discovery of mirror-neurons, is the following:

> **The Fellow-Feeling Paradox:** If fellow-feeling with a principal escalates the original emotion of the principal, we have an anomaly: How could it also attenuate such emotion? Or, *vice versa*, if fellow-feeling with a principal attenuates the original emotion of the principal, we have an anomaly: How could it also escalate emotion?

To express Hume's fellow-feeling paradox formally,

$$E^{d}_{p} = E^{d}_{p}[E_{s}(E^{o}_{p})]$$

whereas E^{d}_{p} is the principal's *derived* emotion; E_{s} the spectator's emotion; E^{o}_{p} the principal's *original* emotion. That is, the principal's original emotion influences the spectator's, which in turn influences the principal's derived emotion.

Fellow-feeling *attenuates* the original emotion when,

$$\partial E^{d}_{p}(E_{s})/\partial E_{s} < 0$$

In contrast, fellow-feeling *escalates* the original emotion when,

$$\partial E^{d}_{p}(E_{s})/\partial E_{s} > 0$$

To rephrase the paradox, how could the same building block of emotion, fellow-feeling, perform two contradictory pathways: "break pedal" and "accelerator pedal"?

As detailed below, Smith focused on the "break pedal" (attenuation) pathway of fellow-feeling. As mentioned above, Hume considered the "accelerator pedal" (escalation) pathway of fellow-feeling to be the usual case. He derided Smith for emphasizing the attenuation pathway,

which he conceded to be illustrated in the attenuation of pain when one comforts a friend upon a loss. Hume also conceded that consolation may actually engender some joy. But Hume, in a ridicule tone, wrote to Smith that if consolation or the attenuation pathway is the usual case, "[a]n Hospital woud be a more entertaining Place than a Ball" [Hume in Smith, 1977, p. 43].

Most economists have focused on the "accelerator pedal" pathway (escalation). For instance, Gary Becker's [1991; 1996; Becker & Murphy, 1993] argues that such social dynamics of consumption is responsible for fads.] theory of social interaction is based on the "accelerator pedal" pathway of emotion/action. The theory shows how particular preference can escalate into a fad or fashion [see also Karni & Schmeidler, 1990]. This escalation pathway can lead fellow-feeling into the wrong path, as in the case mob behavior that resulted in the death of Mr. Morales. But such escalation is the necessary and recommended pathway under a different institutional matrix such as having fun in a ball as Hume noted.

Interestingly, Friedrich Nietzsche condemned Christianity for the same reason. Namely, Christianity promotes "*mitleiden*" (German: *mit*=with, *leiden*=suffering). Unfortunately, the German word "*mitleiden"* is translated into "pity" rather than suffering—given that the term "pity" denotes demeaning status inequality. In any case, Nietzsche's suffering in Christianity is self-indulgence because it is self-centered and, hence, contagious via imitation, i.e., it leads to the escalation of suffering that may push people into lethargy and depression:

> Christianity is called the religion of *pity*. Pity stands in antithesis to the tonic emotions which enhance the energy of the feeling of life: it has a depressive effect. One loses force when one pities. The loss of force which life has already sustained through suffering is increased and multiplied even further by pity. Suffering itself becomes contagious through pity; sometimes it can bring about a collective loss of life and life-energy which stands in an absurd relation to the

> quantum of its cause (--the case of the death of the Nazarene)" [Nietzsche, 2006, p. 488].

Depression, given its contagious character, is at the root of the model of Douglas Bernheim and Oded Stark [1988] concerning what they call "altruism." They reasoned that "nice guys," i.e., altruists, might finish last because no one would want to marry them. Why? Let us say that the partner is depressed. These nice guys would express their sorrow sympathies in a way that escalates the original emotion. This engages, as argued below, the "imitation function." The partners consequently would feel even more depressed.

2. The Mirror-Neuron System (MNS)

The discovery of the mirror neurons is largely attributed to Giacomo Rizzolatti and his co-workers, Vittorio Gallese and Leonardo Fogassi. The amazing central feature of MNS is that it becomes instigated irrespective of whether the spectator undertakes an action, such as grasping an object of significance (cup), or the spectator watches the principal undertaking this action [Rizzolatti *et al.*, 1999; Rizzolatti & Craighero, 2004b; Rizzolatti in Hurley & Chater, 2005].]. The MNS was first discovered in monkeys, located mainly in F5 area of the brain, but later found in dogs and humans:

> Mirror neurons are a particular class of visuomotor neurons, originally discovered in the area F5 of the monkey premotor cortex, that discharge both when the monkey does a particular action and when it observes another individual (monkey or human) doing a similar action [Rizzolatti & Craighero, 2004b, p. 169].

MNS has the following general characteristics:

1. The object of significance to the principal, whether grasping a cup or food, might not be of interest to the spectator—without making a difference as to whether the spectator's MNS is instigated.
2. The principal can be close or far away from the spectator—without making a difference as to

whether the spectator's MNS is instigated.

3. The principal can succeed and be rewarded with the action or can fail—without making a difference as to whether the spectator's MNS is instigated.
4. The principal can be of same species or of totally different species—without making a difference as to whether the spectator's MNS is instigated [Buccino *et al.*, 2004].[6]
5. When MNS discharges, it combines the emotion triggered by the stimulus and the action in response. That is, there is no dichotomy between *emotion* and *action*.

The last point is important to clarify. This paper uses the terms "emotion" and "action" interchangeably and, in fact, recent scholarship on the emotions considers emotion as behaviour [Solomon, 1993]. When the principal becomes angry, he is acting. When the principal acts, it is the continuation of some emotion. Of course, it is often the case that agents feel emotions but tune them down and no body movement takes place. In fact, the spectator's MNS feels/acts with the principal's emotion/action, while the spectator does not replicate the principal's body motion. In these cases, the action is actually inhibited by another neural system that takes into consideration other factors. Given that such factors are not of concern here, there is no need to discuss the inhibitory system and, hence, no need to make further refinement that distinguishes emotion from action.

For our purpose here, the most important feature, stressed by Rizzolatti & Craighero [2004b, p. 170], is that MNS is based on "transitive motion," where MNS is instigated when the spectator observed action moves towards a purpose, such as a hand reaching for a cup. MNS is usually dormant when the spectator observes only "intransitive motion," i.e., action that has no goal or meaning such as the motion of hand with no cup in sight. Such meaningless, intransitive motion

[6] Given MNS operates across nonconspecifics, some institutions can be interpreted as inhibitions. For instance, "halal" (Islamic rule) and "Kosher" (Judaic rule) inhibit the mirror-neuron system, allowing humans to suspend fellow-feeling with animals categorized as food.

does not instigate MNS. But it does instigate another system, called "canonical neurons":

> There are two classes of visuomotor neurons in monkey area F5: canonical neurons, which respond to the presentation of an object, and mirror neurons, which respond when the monkey sees object-directed action. In order to be triggered by visual stimuli, mirror neurons require an interaction between a biological effector (hand or mouth) and an object. The sight of an object alone, of an agent mimicking an action, or of an individual making intransitive (nonobject-directed) gestures are all ineffective [Rizzolatti & Craighero, 2004b, p. 170].

Broadly speaking, a spectator's MNS is instigated when the spectator focuses on the intention behind the principal's action, i.e., *understands* or exercises so-called "theory of mind" or "mindreading" [Meltzoff & Gopnik, 1993; Baron-Cohen, 1995; Baron-Cohen *et al*., 2000]. In contrast, a spectator's CNS is instigated when the spectator does not focus on the intention behind the principal's action, i.e., understanding ("theory of mind") is irrelevant. So, CNS is stimulated simply by observing an action while totally detaching from the cause (incentive) behind it.

Whether imitation also involves "theory of mind" (understanding) is still an open question [see Tomasello, 1999; Tomasello *et al.*, 2005]. But it is the central thesis of this paper that imitation does not involve understanding. This does not mean that an organism which is incapable of understanding can be still capable of imitation. Actually, as Rizzolatti and Craighero [2004b] report, CNS (imitation) is evident in probably very limited upper primate species, but definitely in humans, while most primate species share the MNS (understanding). So, acts of imitation undertaken by humans must already mean that humans are capable of understanding. But this finding does not mean that "to imitate" is the same as "to understand."

The literature so far does not make a clear differentiation between MNS and CNS [see

Hurley & Chater, 2005, vol. 1, ch. 1]. Rizzolatti and Craighero [2004b] even argue that MNS is involved in both functions, understanding and imitation. They maintain that language acquisition is greatly based on imitation, where the spectator (child) mimics the adults (principals) without understanding.

While the two functions, viz., understanding and imitation, somewhat overlap, it is important to distinguish them. With imitation, there is no understanding. With understanding, there is no imitation, which gives rise to what is dubbed here the "mirror-neuron paradox":

> **The Mirror-Neuron Paradox:** If mirroring the action of a principal is imitation, we have an anomaly: How could it be distinguished from understanding? Or, *vice versa*, if mirroring the action of a principal is understanding, we have an anomaly: How could be distinguished from imitation?

It is insufficient to trace the imitation-understanding functions back to some neural substrate. This would only postpone the question: How could one neural substrate be invoked in one context, but remain dormant in another?

3. Smith's Solution of the Paradox

Smith offered a solution to Hume's objection. The solution should prove to be useful in one respect: It clearly shows us how the mirror-neuron paradox is basically Hume's fellow-feeling paradox.

Smith defines sympathy as understanding which involves approval of propriety. As such, sympathy can explain, on one hand, the attenuation pathway of fellow-feeling and, on the other, the "understanding function" of the neural system. A few others, as noted above, have noted the connection between Smith's notion of sympathy and the "understanding function" of MNS [see Rustichini, 2005; Rizzolatti & Craighero, 2004a].

The first paragraphs of *The Theory of Moral Sentiments* betray the fact that Smith [1976] took Hume's paradox very seriously. The paragraphs show Smith's major theoretical innovation concerning sympathy. What matters for sympathy is understanding or the "good sense" mentioned earlier with respect to the death of Morales: When the spectator expresses sympathy, the spectator *is* considering the incentive (stimulus) that occasions the action/emotion of the joyful or distraught principal. The consideration of the incentive, as the necessary condition of sympathy, is responsible for the attenuation, "break pedal" function of fellow-feeling.

As detailed elsewhere [Khalil, 2007e], the "break pedal" function of sympathy in Smith is widely recognized [see Haakonssen, 2002, p. xiv; Levy & Peart, 2004, p. 334, n. 3; Fontaine, 1997; Gordon, 1995]. Nonetheless, it is important to identify its features to see why Smith's solution is only partially successful.

3.1 Attenuation of Emotion arises from Understanding

To answer Hume's objection, Smith posited a particular definition of sympathy. For sympathy to take place, the spectator must transport himself or herself into the station of the principal. The spectator must try to enter the station of the principal and examine his or her emotion/action in relation to the incentive. If such transportation is successful, the spectator's sympathy is nothing other than the recognition that the principal's action is rational. A principal takes a rational action when he or she executes the optimal decision, i.e., the one that maximizes *ex ante* utility. When the spectator puts himself in the shoes of the principal, the spectator is assessing whether the action is proportional to the incentive, i.e., rational. So, the issue of rationality is at the core of Smith's

sympathy.

To elaborate, to solve Hume's paradox, Smith redefines and narrows the meaning of the term "sympathy." Sympathy is not the contagious emotion suggested by imitation. Rather, it is about understanding. As such, one should not be perplexed, as the case with Hume, when sympathy attenuates the original emotion.

To wit, Smith's concept of sympathy-as-understanding corresponds well with the "understanding function" of MNS. Namely, in its function as understanding, the spectator's MNS becomes instigated only when the spectator observes that the principal is involved in transitive action, i.e., objected-directed action or action in relation to stimulus (incentive). So, the spectator does not simply imitate the emotion/action of the principal. The spectator can only replicate the principal's emotions if such emotions are understood, i.e., how is the action related to the stimulus.

To elaborate, the spectator who is trying to understand the emotions of the principal is not any spectator. Rather, he or she is an "impartial spectator." Smith's concept of the "impartial spectator" and the mechanism that gives rise to what he calls "propriety" is involved [Khalil, 1990, 2006]. Stated briefly, as an impartial agent, the spectator is, by definition, a judge of whether the principal's action/emotion is proportional to the stimulus (incentive). That is, the judge has to determine if the action is optimal given the incentive. When the principal also acts as the judge, then the impartial spectator resides within the principal, as second self, and called throughout the "judge within." The judge within, to note very briefly, is *not the internalization* of social norms à la functionalist sociological theory or microsociological theory à la George Herbert Mead [see Khalil, 1990]. Rather, for Smith, the judge within is simply the principal looking after the utility of the

current self as well as the utility of the future self—i.e., examining the demands of each self from a distance that is occupied by the impartial spectator as well. So, there is a hall of mirrors in Smith's analysis—where there is a correspondence between the judge within and the judge without (impartial spectator). But it is important not to get lost in such a hall and keep our focus on the "original copy." For Smith, and in this paper, the "original copy" is the judge within, which is usually *externalized* in case our agent is not a Robinson Crusoe. The original copy is not the judge without or what agents think is the judge without, as implied in the analysis of Jean-Pierre Dupuy [2004, 2006].

In this analysis, the principal, to be rational, takes action in light of the choice determined by the judge within. This interpretation of Smith's theory makes it a theory about "self-command," which is one of the main virtues of Smith's book. When the principal exercises self-command, the principal is accommodating the needs of the current self which, given the scarcity of resources, competes with the needs of future self. And the principle, or judge within, is capable of doing such an accommodation by examining the needs of the current self from a distance, which Smith calls metaphorically the "impartial spectator." That is, the impartial spectator is nothing but a metaphor for non-myopic decision making, where the utility of the future self must count. So, there is no fissure between the judgment of the caring, impartial spectator and the judgment of the self concerning its own welfare--a fissure that is supposedly deep and need to be bridged according to Stephen Darwall [2002, 2006].

So, Smith's theory of sympathy is ultimately about rational intertemporal allocation when time inconsistency (temptation) is a problem. Smith's theory anticipates the dual-self model of

intertemporal choice that is gaining attention not only in economics [Thaler & Shefrin, 1981; Fudenberg & Levine, 2006] but also in biology with regard to intrapersonal conflicts [e.g., Haig, 1993, 2003; Burt & Trivers, 2006]. This might come as a disappointment for the new scholarship on Smith, such as Deirdre McCloskey's [2006], which promotes Smith as an alternative to narrow standard theory of choice [e.g., Gintis *et al.,* 2005]. Smith's concepts of sympathy and self-command are ultimately about optimal choice.

However, Smith's theory is not that simple. It provides a rich account of the mechanics of self-command, i.e., how exactly does the self enforce time consistency and succeed in fighting temptations? Here, Smith divides the agent into principal and impartial spectator, which is expressed externally as "doer" and "judge," respectively [see Khalil, 1990]. If the principal (doer) becomes very agitated as a result of a simple failure, and surrenders to anger, the impartial spectator (judge) would not sympathize with the principal. This means that the impartial spectator or, for short, the spectator cannot approve the principal's action/emotion. For the principal to win the approbation of the spectator, the principal must take residence in the spectator's station, i.e., look at his current anger from a distance. Such an examination would allow him to see that if he acts with anger, he might hurt future self. So, a judge has to restrain current self so that the future self is not hurt. But how does this exactly work? For Smith, the current self seeks the sympathy of the judge. The judge, acting, as a spectator, cannot provide sympathy, i.e., approval about the efficiency of the action, if the pitch of action/emotion of the principal is too high or disproportional to the cause (incentive). The principal, hence, must lower the pitch of emotion/action to win the approbation, i.e., sympathy, of the spectator. So, the act of sympathy can be interpreted as nothing but the fact that the judge is

taking into consideration the interest of future self as well.

If the principal lowers the pitch of emotion/action, it would be easier for the spectator to travel and enter, i.e., sympathize, the station of the principal. As long as the principal is too angry or too joyful relative to the incentive, the impartial spectator simply cannot understand the emotion/action of the principal, i.e., approve. So, for Smith, the spectator's understanding (empathy) automatically entails approbation, i.e., judgment of propriety. Smith's notion of "sympathy" is nothing but the collapse of empathy (understanding), on one hand, and propriety (efficiency of action), on the other. So, sympathy is nothing but the conclusion that the principal has acted with propriety (optimally), which allowed the spectator to understand it.

Furthermore, Smith argues that when the principal acts properly, he or she also experiences "another source of satisfaction" [Smith, 1976, p. 14; see also p. 209]. It is the satisfaction of being proper and making good choices—which amounts to self-regarding, self-worth, or self-esteem. So, when the principal in pain receives a consolation from a friend, it amounts to approval that the principal acted efficiently or properly. Such an approval occasions a sense of joy, i.e., the sense of self-worth, as when one succeeds in taking any efficient action. Smith argued, against Hume's utilitarianism, that such self-regarding cannot be reduced to the usual utility. In any case, self-regarding is squarely based on rational action or propriety [Khalil, 2007e,f]. This cannot be detailed here. But it suffices to add that such self-regarding or self-worth plays a crucial role in the ranking of worth of different persons, which is at the origin of status inequality, social segregation, and adulation as discussed below.

3.2 Why Smith's Solution Fails

The question here is whether Smith succeeded in answering Hume's objection. Obviously, Smith has succeeded in identifying the operation of sympathy-as-understanding that leads to the attenuation of original emotion. But did Smith explain the coexistence of the attenuation-escalation pathways? Or, to put it in a modern dress, did Smith explain the coexistence of the understanding-imitation functions, the origin of the mirror-neuron paradox?

Smith failed to show how the same primitive emotion, fellow-feeling, can also give origin to the two pathways highlighted in Hume's letter. To put it in the terms of the mirror-neuron paradox, Smith showed the roots of the "understanding function" of fellow-feeling, which is behind attenuation of emotion. Smith does not reconcile the "understanding function" with the "imitation function" of fellow-feeling, which is behind the escalation of emotion.

This does not mean that Smith was unaware or ignorant of the "imitation function" of fellow feeling. In fact, Smith recognizes it when he discusses, e.g., the pleasure of company when people read a book together as opposed to reading it alone [Smith, 1976, p. 14]. As Martin Hollis [1998] notes, it is usually pleasurable to converse with people who had similar experiences. When one reads a book, watches a film, purchases a new automobile, or dines at a restaurant, it would be more pleasurable to converse with others who had undergone the same experience. Such conversation enhances the marginal utility as a result of the escalation effect, as discussed earlier with respect to Gary Becker's theory of social dynamics and fashion.

To recognize the "imitation function" of fellow-feeling, which is responsible for the escalation pathway, is one matter. It is another matter to show how the same fellow-feeling can

occasion escalation in one case and attenuation in another. Smith failed to show such double function of fellow-feeling. Thus, he failed to resolve the mirror-neuron of Hume.

4. The Two-Axis Evaluation Hypothesis (TAE)

4.1 What is Wrong with Smith's Solution?

How to account for the co-existence of attenuation-escalation pathways? As a necessary step, but not a sufficient one, this paper questions Smith's thesis that the "understanding function" of fellow-feeling *necessarily* entails approbation concerning propriety, and *vice versa*. Smith felt that he needed to make such an assumption to account for why sympathy, e.g., with a grieving parent does not escalate the original grief. The attenuation can be explained by the approbation of propriety alone—i.e., without assuming that the spectator "understands" the intention of the principal. In fact, we need to set aside the "understanding function" from the approbation of propriety in order to explain the escalation pathway.

From casual empiricism, it is often the case that spectators express understanding of the agony of someone—while withholding approval. For instance, many supporters of Israel understood Israel's little 2006 summer war, in which it killed over one thousand Lebanese, over 90% civilian, in retaliation for Hezbollah's capture of two Israeli soldiers. But many of these supporters of Israel did not approve of Israel's little war in the sense of finding it to be well-calibrated to the cause. That is, they did not find Israel's action proper or efficient in serving Israel's long-term interests.

Likewise, the spectator may understand Harry Truman's decision to drop the nuclear bombs on the two Japanese cities full of civilians. Or, one may understand the US-UK invasion of Iraq in

March 2003. But does such understanding entail approbation of propriety? Were such actions *ex ante* rational or did they amount to succumbing to anger, temptation, or recklessness in response to fleeting and enticing opportunity? One may understand that one, under the temptation of superior power or the temptation of a desert tray, succumbs to weakness of will and acts myopically. But such understanding does not entail approval of propriety.

So, approbation concerning the rationality of action is a question about whether the action is proper (optimal), while approbation concerning the intentionality of action is a question about whether one can understand (empathize with) the action. For instance, one may have a commitment to restrain himself even if one has momentary military superiority or instantaneous confrontation with a desert tray. And to act contrary to either commitment, i.e., react proportionally to the stimulus, makes us judge the action as improper or what economists call "inefficient." But such judgment does not entail that we failed to understand the principal's utility and constraint functions. If we judge that Truman *ex ante* acted improperly (inefficiently) when he approved the dropping of nuclear weapons on civilians, it does not mean we do not understand why he did so. Truman's intention is to enhance the wellbeing of American soldiers, i.e., to bring a speedy conclusion to the conflict that may, one can argue, save also Japanese lives.

In this light, the act of understanding (empathy) need not entail approbation of propriety, i.e., judgment concerning the rationality of the act. And *vice versa*, the judgment concerning rationality does not entail empathy. For instance, we can be impressed with the efficiency of a serial killer, the Nazi Holocaust organizer, or a cult leader. But this does not entail that we understand, in the sense of empathize, with the intention of the agent.

4.2 Why Is Intentionality Outside Economics?

Contrary to Smith, this paper does not collapse the axes of intentionality judgment and efficiency judgment. This paper conjectures that the two axes are orthogonal. Let us call the axis that invokes the efficiency judgment the "rationality axis," while call the axis that invokes the understanding judgment the "intentionality axis." The rationality axis occasions the familiar judgment arising from the question that economists usually ask: Is the action proper (efficient) or improper (suboptimal)? The intentionality axis occasions a less familiar judgment at least for economists: Is the action understandable or is it revolting?

First of all, what does the term "understandable" mean? The term is used in this paper in the sense that the spectator finds the action amiable, i.e., it does not arouse disgust as when a North American is disgusted when he or she sees someone eating cockroaches or snakes [Rozin *et al.*, 1993]. The spectator would "empathize" with (understand) amiable actions while the spectator would be disgusted with revolting actions. In this sense, the term "understanding" (i.e., empathy) should not be confused with the term "comprehension"—where comprehension is defined as the scientific account of why hurricanes, earthquakes, wars, genocides, cults, sunspots, and serial killing take place. Comprehension, or whatever term one chooses to denote scientific accounts, has nothing to do with the evaluation of the end—i.e., whether the end is revolting or understandable. A few economists confused "understanding" (i.e., empathy) with "comprehension"—which is responsible for an unfortunate mess [e.g., Binmore, 1998; Harsanyi, 1977].[7] This paper maintains the original

[7] Ken Binmore uses the term "empathy" in the sense of comprehension when he describes how a

meaning of "empathy."[8] The original meaning is not about comprehension. Put tersely, to comprehend hurricanes and hate crime does not entail empathy. Empathy or understanding is rather a judgment about the end or the *intention* of the principal. (To wit, hurricanes can become also the subject of "understanding" if one believes that hurricanes are intended by gods or demons.)

The TAE hypothesis stipulates that judgment about the intention of an action differs from a judgment about the rationality of an action. The intentionality judgment is outside the scope of economics. The intentionality axis does not involve allocation of resources. For resources to be allocated, and hence to be within the scope of economics, the resources must be, first, scarce, and, second, fungible. Along the intentionality axis, resources seem neither nor scarce. The principal's intention, whether hopeless or hopeful, is a matter of perspective.

The issue of perspective is involved and cannot be discussed in detail here. It is the bread-

---

gunfighter wants to know the position of an opponent:

> Adam sympathizes with Eve when he so identifies with her aims that her welfare appears as an argument in his utility function. ... The extreme example is the love a mother has for her baby. Adam empathizes with Eve when he puts himself in her position to see things from her point of view. Empathy is not the same as sympathy because Adam can identify with Eve without caring for her at all. For example, a gunfighter may use his empathetic powers to predict an opponent's next move without losing the urge to kill him [Binmore, 1998, p. 12].

Also Harsanyi [1977] uses the term "empathy" in the sense of comprehension and assessment of position of others (opponents or loved ones). Harsanyi distinguishes empathy from "subjective preferences" or what Binmore [1994, 1998] and Amartya Sen [1977] call "sympathy." Psychologists, such as Michael Basch [1983], also use the term "empathy" in the sense of comprehension.

[8] According to Gerald Gladstein [1984, p. 40; see also Gladstein, 1987], the term "empathy" was coined in 1909 as a translation of the German *einfühlung* (from ein "in" + fühlung "feeling"). The German word, popularized by Theodor Lipps [1960], was coined in 1858 by German philosopher Rudolf Lotze (1817-81) from the Greek *empatheia* "passion," from en- "in" + pathos "feeling."

and-butter the framing effect as starkly highlighted in the famous Asian disease experiment [Tversky & Kahneman, 1981]. The experiment shows that one's perspective, i.e., how to view an event, influences one's action. One is usually ready to act as a risk-averse agent if one examines an event via the gain frame (optimism), while one is normally ready to act as a risk-loving agent if one examines an event via the loss gain (pessimism). One has to process information by one frame or another. The frames are not fungible. They are also costless because the switch from one frame to another is a matter of perspective. To see a glass as half-empty or to see it as half-full, does not involve expenditure of resources [Khalil, 2007d].

If so, the intentionality axis is outside the scope of resource allocation. It is rather about being hopeful—and as such it is about the pursuit of ends that are supportive of life—as opposed to being hopeless. If the principal lacks hope, the principal would take views or even actions that undermine life. The spectator would not empathize with such actions; the spectator would rather find them revolting. Put differently, the intentionality axis is about the "choice" to pursue life as opposed to quitting life altogether. It one decides to opt out, through self-sabotage, it is not a good "chosen" in the conventional sense in economics. It is not a good to start with.

So, self-sabotage in the form of addiction, suicide, and disgusting ends that undermine one's welfare must be motivated by hopelessness, hate, envy, or malice. The emotion of envy, as discussed below, is directly related to schadenfreude and evil. Maury Silver and John Sabini [1978, p. 108] measure how people perceive envy. They find it to be an emotion that does not involve any benefit—sinful or non-sinful or illicit or non-illicit. Envy may not involve any bodily action or

motion. It can be simply the emotional difficulty of congratulating colleagues upon their success. So, envy has no opportunity cost; allocation of resources is not involved. The intentionality axis does not involve decisions about allocation of resources—such as whether should one live in neighborhood A as opposed to neighborhood B, with all the tradeoffs of benefits and costs. It is rather about deciding whether to enhance living or to succumb to hopelessness in the forms of malice, sabotage, or self-sabotage that is usually called addiction.

For economics, it is an odd question to ask if one prefers to hurt themselves or reduce wellbeing: Why would anyone choose to reduce wellbeing (via sabotage and self-sabotage) for no apparent or hidden benefit whatsoever? But such an odd question is exactly at the core of hope, malice, internal motivation, and so on. This "odd" question is the bread-and-butter of the behavioral sciences.

When the intentionality axis is engaged, the spectator *puts himself in the shoes of the principal* and asks whether the action supports wellbeing (and hence understandable) or aimed only at reducing wellbeing as a result of envy or hopelessness (and hence revolting). To note, when the intentionality action is disengaged, the spectator does not ask such question. He rather experiences stimulus while *remaining in his or her own shoes*. That is, the spectator would be self-centered, not interested in understanding the action of the principal in relation to whether it supports wellbeing or not. The spectator would simply use the action of the principal as a stimulus for his or her own senses or memories.

When the intentionality axis is engaged, the spectator is trying to evaluate whether he can empathize with, i.e., understand, the principal's action in relation to the intention. If the spectator

finds the principal's action is intended to enhance wellbeing, but may or may not be the optimal action, the spectator still would empathize with it. Otherwise, if the spectator finds the principal's action is intended to undercut wellbeing for its own sake, the spectator would find it revolting or disgusting.

4.3 What is Disgust?

But what is disgust to start with? It is a complex emotion that has a physiological component [Miller, 1997]. As defined by P. Rozin and A.E. Fallon, disgust is the "[r]evulsion at the prospect of (oral) incorporation of an offensive object. The offensive objects are contaminants; that is, if they even briefly contact an acceptable food, they tend to render that food unacceptable [quoted in Rozin *et al.,* 1993, p. 576]. At face value at least, when a spectator finds an item of consumption, such as cockroaches or snakes, revolting, it is *usually* accompanied by the belief that the item is dirty, carrier of diseases, or simply detrimental of wellbeing. Our spectator may even experience nausea and sickness in the stomach when he watches the principal eating cockroaches or snakes. Thus, revulsion is based on the belief concerning wellbeing.

> **Definition:** The spectator judges an emotion/action as revolting or disgusting when he watches a principal taking such action that undermines wellbeing while it does not promote the wellbeing of *anyone*.

Given this definition, *revulsion* differs from *squeamishness*. While revulsion is about wellbeing, squeamishness arises from associative memory rather than the result of beliefs concerning wellbeing. The spectator may not drink orange juice, maybe for many days or forever, if it happened that the principal who ate cockroaches also drank orange juice. Similarly, a squeamish spectator

may not eat chicken, maybe for many days or forever, after a visit to the slaughterhouse. The slaughter and the associated blood evoke self-centered memories and associative feelings, i.e., indulgence as discussed below. It might be difficult empirically to distinguish revulsion from squeamishness. But at least theoretically, there must be an original revulsion, concerning wellbeing, in order for associated constructs, i.e., squeamishness, to arise.

Further, given this definition, *revulsion* differs from *suboptimality*—following the fault dividing the intentionality and the rationality axes. The spectator judges an action as revolting by *judging the end itself*: does it undercut wellbeing without the intention, and even without the intention of the natural selection mechanism, to enhance the wellbeing of *anyone*? In contrast, the spectator judges an action as suboptimal by, first, taking the end as given and, second, by *judging the means*: do such means achieve the end in the most efficient manner? That is, in judging optimality, the spectator judges only whether the action, such as favoring the self, is the best method to advance such end—while taking the end as given. In some cases, or actually in most cases, the end is not revolting. But still the action can be deemed as suboptimal when the principal is myopic, i.e., the principal advances the utility of the present self but at an unjustifiable cost to future self.

Such suboptimality is noted by psychologists who study disgust, viz., Paul Rozin *et al*. [1993], and they call it "rights violations." They note that there is a difference between disgust and rights violations (suboptimality). But they fail to locate the source of the difference, which would be the outcome of failing to distinguish the rationality axis from the intentionality axis.

4.4 The Crankcase Oil Problem

So, revulsion or disgust is based on the belief that the action is detrimental to wellbeing, i.e., an evaluation along the intentionality axis. The belief is not about the rationality axis, i.e., the extent to which it is the *best way* to support an end—which can be any end. The belief is rather about a binary judgment of whether the action is aimed to support or to sabotage a particular end, viz., wellbeing.

Put differently, let us say that our same spectator, who is revolted by cockroaches or snakes, volunteers and eat them for no apparent or hidden compensation. How can we model such an action that is aimed at the reduction of wellbeing? It would be *ad hoc* to state that the spectator must have changed his or her mind and now prefers the disgusting item. It amounts to "sticking new tastes in the utility function." As George Stigler and Gary Becker [1977] have warned, this practice leads to the instability of the utility function [see Khalil, 2008]. Similarly, one cannot say that someone must love crankcase oil if he or she starts dinking it, what Robert Frank [2006, p. 231] calls the "`crankcase oil' problem."

To elaborate, the crankcase oil problem is based on George Stigler's famous quip: How should we model a person who drinks crankcase oil from his automobile while fully knowing that it is neither medicinal nor tasty, but it is rather detrimental to wellbeing? If we assume that the person simply likes the crankcase oil, it would violate the principle of stable preferences. We simply cannot move item Z from the category of "garbage" to "goods," i.e., assume that preferences have changed, simply because the agent started to consume Z.

If we maintain the standard position, viz., the rationality axis is the only axis of evaluation, we face the anomaly of the crankcase oil problem. That is, we would be unable to explain revolting or destructive behavior. The drinking of crankcase oil, or having revolting intention to lower

wellbeing for no one's benefit, is not an issue about prices and budget constraints where the rationality axis would be relevant. The issue of drinking crankcase oil is rather an issue about choosing to survive or not survive—an issue that does not involve a question about allocation of resources.

4.5 Harry Truman and Serial Killers, Again

Given that the intentionality axis cannot be reduced to the rationality axis—viz., the TAE hypothesis—we can pinpoint exactly the difference in our evaluation of the actions of Harry Truman and the serial killer. We use the rationality axis when we find that the serial killer has, e.g., acted *ex ante* irrationally—because the killer succumbed to opportunities that were *ex ante* clear to be suboptimal. Likewise, the spectator using the rationality evaluation might find that Truman has succumbed to myopic benefits—because it makes it excusable for other countries in future conflicts to drop nuclear bombs on US cities. But neither judgment entails any conclusion concerning the intentionality evaluation. Namely, the spectator using the intentionality evaluation might find the killer's action revolting while Truman's action understandable. But, of course, Truman's action would also be evaluated as revolting if one finds that Truman was motivated by hatred and spite.

But in order to make such fine distinction, we need to distinguish the rationality axis from the intentionality axis, the core of TAE hypothesis.

4.6 When the Two Axes are Disengaged

The mirror-neuron paradox is not solved yet. While Smith's account of approbation can explain the

attenuation pathway, we still have to explain the escalation pathway. To do so, we took the necessary step above, i.e., the separation of the intentionality axis from the rationality axis. We now need to take the sufficient step, i.e., examine the processing of fellow-feeling when either of the two axes is actually *dis*engaged.

But how could an axis be disengaged? Let us first map the structure of possibilities: The rationality axis can be disengaged while the intentionality axis is engaged, and *vice versa*. Or both can be disengaged. Or both can be engaged. We have four possibilities, as Figure 1 shows.

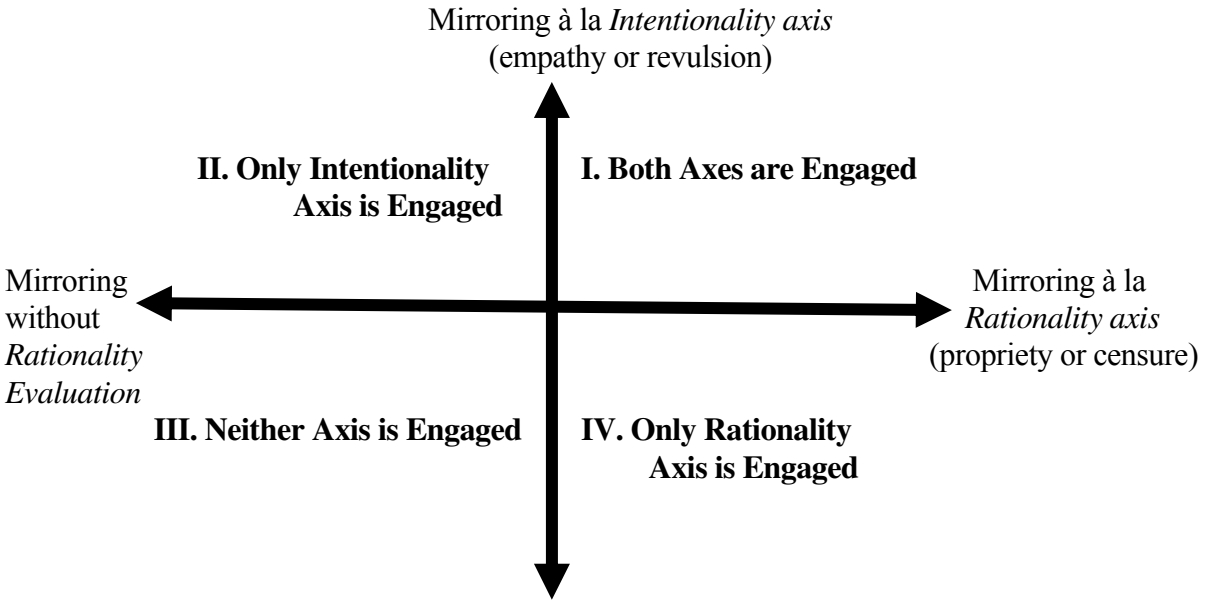

Figure 1: The Two-Axis Evaluation Hypothesis (TAE)

Quadrant I shows the combination when both axes are engaged, i.e., when the spectator is involved in efficiency evaluation and the attempt to understand the intention. Quadrant II demonstrates the combination when only the intentionality axis is engaged, i.e., when the spectator is involved exclusively in the attempt to understand the intention. Quadrant III displays the combination when neither axis is engaged, i.e., the spectator is involved exclusively in imitation, where intentionality and rationality of the principal's action are ignored. Quadrant IV exhibits the combination when only the rationality axis is engaged, i.e., when the spectator is involved in imitation while trying to evaluate the rationality of action.

The "understanding function" and the "imitation function," the core of the mirror-neuron paradox, can be pure or modified. Either one can be pure, as afforded in either quadrants II or quadrant III. In either of these quadrants, there is no judgment of efficiency. But with the judgment

of efficiency, when rationality axis is engaged, we have qualified understanding (quadrant I) and qualified imitation (quadrants IV).

The issue of disengagement is paramount if we want to understand the attenuation-escalation pathways. It should become clear that the attenuation pathway is the outcome of engaging the rationality axis. In contrast, the escalation pathway is the outcome of disengaging both, the rationality and the intentionality axes. From common sense, one would tend to exercise self-command when one is thinking from a distance about the most effective or proper action. However, if one suspends such judgment, and especially when one also suspends the intentionality axis, imitation takes over which feeds on itself, leading to escalation.

Now, to engage an axis, one needs only to ask the relevant question. If the rationality axis is engaged, one asks: can one approve of the rationality of the action/emotion or is it suboptimal? If the intentionality axis is engaged, one asks: can one empathize with the action or is it revolting? But what does it mean to have an axis disengage? For the rationality axis to be disengaged, one does not judge whether it is rational or not. The fellow-feeling or mirroring takes place without such assessment. For instance, if a serial killer commits an *ex ante* irrational act and gets caught, one may suspend the rationality axis, and only engage the intentionality axis (quadrant II): So, one would not judge the rationality of the act. One would only ask whether the act is understandable. And of course, it is not understandable. Likewise, if a man drives fast above the speed limit because he is late to a concert, and loses control over his car and hits a crowd of people and kills a dozen of them, one may suspend the rationality axis and only engage the intentionality axis (quadrant II): is the action of the young man understandable? And of course, it is understandable.

When one asks whether an action understandable—i.e., can one empathize with the agent—one is examining the principal's behavior in relation to the principal's intention. We are not examining behavior in relation to incentives—which would be a question along the rationality axis. For instance, the serial killer might have killed in total, before being caught, a half dozen people. But his intention would be examined differently from the driver who killed a dozen out of recklessness.

Now, what if the intention is not considered at all? Here, the spectator processes the fellow-feeling while the intentionality axis is suspended. The spectator only senses the action without the intent, i.e., without examining whether it is motivated by wellbeing or by envy. But such an observation, if it registers emotion in the spectator, the emotion must be the result of re-consuming or remembering past experiences. So, the emotion of the spectator has little to do with the situation. The situation is not even the subject of understanding or no-understanding. Rather, he spectator, involved in his own station or circumstances, uses the stimulus or observed action, to re-call how he would feel if the observed event happened to him.

The disengagement of the intentionality axis actually informs ego-centric theories of altruism stretching from Thomas Hobbes to Gary Becker [see Khalil, 2001, 2002b, 2004]. These theories, known also as "warm-glow" theories, the spectator/benefactor contributes to the wellbeing of the principal only insofar the excitement or utility of the principal excites, in reflection, the excitement of the spectator/benefactor. Here, the benefactor does not care about the intention of the principal. The benefactor is only interested in how the excitement of the principal enhances his own utility.

Such a view of altruism does not distinguish between altruism and social interaction behind

the rise of fads and escalation of fashion. Gary Becker [1996] lumps both phenomena almost under the same model of social interaction.

To wit, as alluded above, fads resemble the escalation of original feeling, the basis of Hume's sympathy-as-imitation. Here, the original emotion is amplified, and original action is extended, as others imitate the principal's action. The principal starts to reap greater marginal utility as others enact the same fashion or become in-synch with his mood. In such a situation, others imitate the principal without attention to his intention.

So, escalation of original emotion takes place when the intentionality axis is disengaged. Such escalation need not involve judgment of propriety. In Hume's example, quoted above, a merry person makes other merry, via contagion, where others do not pass judgment on the rationality of the mood. To wit, to ensure the contagious aspect of fads or moods, agents do not invoke the rationality axis.

So, the primitive fellow-feeling gives rise to escalation when the two axes are disengaged, which is depicted as quadrant III. The same primitive can give rise to the attenuation of emotion if the two axes are engaged, which is demonstrated as quadrant I. In quadrant I, even if the act is revolting—such as genocide or mass killing motivated by hate—it can still be judged according to the rationality axis. While one cannot empathize with such an act, one can still judge its efficiency. And such judgment of efficiency entails that the serial killer must not take short-cuts or given in to excitement and anger, if he does not want to be caught.

While Smith's concept of sympathy can also, as shown earlier, explain attenuation of original fellow-feeling, it is limited in scope. It cannot explain attenuation in cases when

understanding is impossible, such as in serial killing, while rationality is possible. Smith's analysis, given its conflation of understanding with propriety, lacked the analytical tools to account for wider phenomena of propriety when understanding is lacking.

Of more importance, given Smith's conflation of the two axes into one, and not realizing the consequences of suspending approbation, Smith's analytical tools cannot capture the four quadrants just discussed. Therefore, Smith's analysis of fellow-feeling cannot explain how the same primitive can lead to escalation of original emotion, and not only to its attenuation.

So, the proposed TAE hypothesis solves the mirror-neuron paradox. The same primitive, fellow-feeling, can lead to the attenuation or escalation of original fellow-feeling. This depends on whether both axes are engaged, which would lead to attenuation, or whether both axes are disengaged, which would lead to escalation.

Furthermore, the TAE hypothesis sheds brighter light on the two functions of MNS and CNS discussed earlier, viz., the "understanding function" and the "imitation function" of mirroring. When the intentionality axis is engaged, the "understanding function" or, in case of revulsion, disgust, is operative. When the same axis is disengaged, there is neither understanding nor disgust. The judgment concerning intentionality is totally shelved or frozen. In such case, the "imitation function" is operative. So, the two functions are not incompatible. The functions diverge simply because the primitive fellow-feeling is processed along different institution or different part to the intentionality axis.

5. Testing the TAE Hypothesis

5.1 Testing the Rationality Axis

To test the rationality axis, we can set up the following benchmark:

1. Spectators observe principals who are stimulated by incentives of different intensity (winning 1 banana to a box of fruits).
2. Records are kept of the action/emotion of principals and the corresponding spectators' MNS.
3. Principals are aware of the fact that they are being observed, but do not know the nature of the experiment.

As for the treatment,

1. Repeat steps #1-2 above
2. Principals are aware of the nature of the experiment, and their reactions are no longer of their choice. Rather their reactions are selected for them by the experimenter so that they widely differ from the benchmark case. As for the spectators, they are not informed that the reactions of the principals are manipulated.

The TAE hypothesis predicts the following. As for the intentionality axis, the spectators' canonical-neuron system (CNS) is irrelevant: it should be the same in the benchmark as in the treatment. In both cases, there is an approval of the intentionality of principals' action since the fruits are seen to be conducive to wellbeing. The focus here is rather on MNS. If it is engaged, the spectators' MNS should behave differently in the treatment case. It should reflect impropriety. If it is not engaged, the spectators' MNS should not register any activity.

5.2 Testing the Intentionality Axis

To test the intentionality axis, it is more problematic because the wellbeing of principals cannot be harmed. Nonetheless, the harm can be measured without actually inflicting harm on the principals as shown in the treatment.

Let us start with the following benchmark:

1. Spectators observe principals who are eating "culturally understood" desert (such as most fancy ice cream with strawberry topping).
2. Records are kept of the action/emotion of principals and the corresponding spectators' MNS.
3. Principals are aware of the fact that they are being observed, but do not know the nature of the experiment.

As for the treatment,

1. Spectators observe principals who are eating "culturally disgusting" desert that is clearly knowable to the spectators (e.g., fancy ice cream with chopped liver topping).
2. Records are kept of the action/emotion of principals and the corresponding spectators' MNS.
3. Principals are aware of the nature of the experiment, and their reactions are no longer of their choice. Rather their reactions are selected for them by the experimenter so that they exhibit the usual emotions/excitement as if they are eating "culturally understood" desert. As for the spectators, they are not informed that the reactions of the principals are manipulated.

The TAE hypothesis predicts the following. As for the rationality axis, the spectators' MNS should be the same in the benchmark as in the treatment. In both cases, there is an approval of the propriety of the action of the principals. The focus here is rather on the CNS. If it is engaged, the spectators' canonical-neuron system would behave differently in the treatment case. It should reflect revulsion or absence of empathy because the food is judged as a hindrance to wellbeing. It is similar to an act of hurting one's own body since revulsion arises from the belief that the action reduces even momentary wellbeing. If CNS is not engaged, the spectators' CNS should experience same excitement in the treatment as in the benchmark. The spectators' CNS would imitate the apparent excitement of the principals.

6. Four Families of Fellow-Feeling

What is payoff of TAE hypothesis? It allows us to differentiate among four "families" of fellow-feeling. If nothing else, this should allow us to solve the mirror-neuron paradox in the sense of

identifying the "understanding function" and the "imitation function" and showing how the attenuation-escalation pathways can co-exist. As Figure 2 shows, each cluster belongs to a different

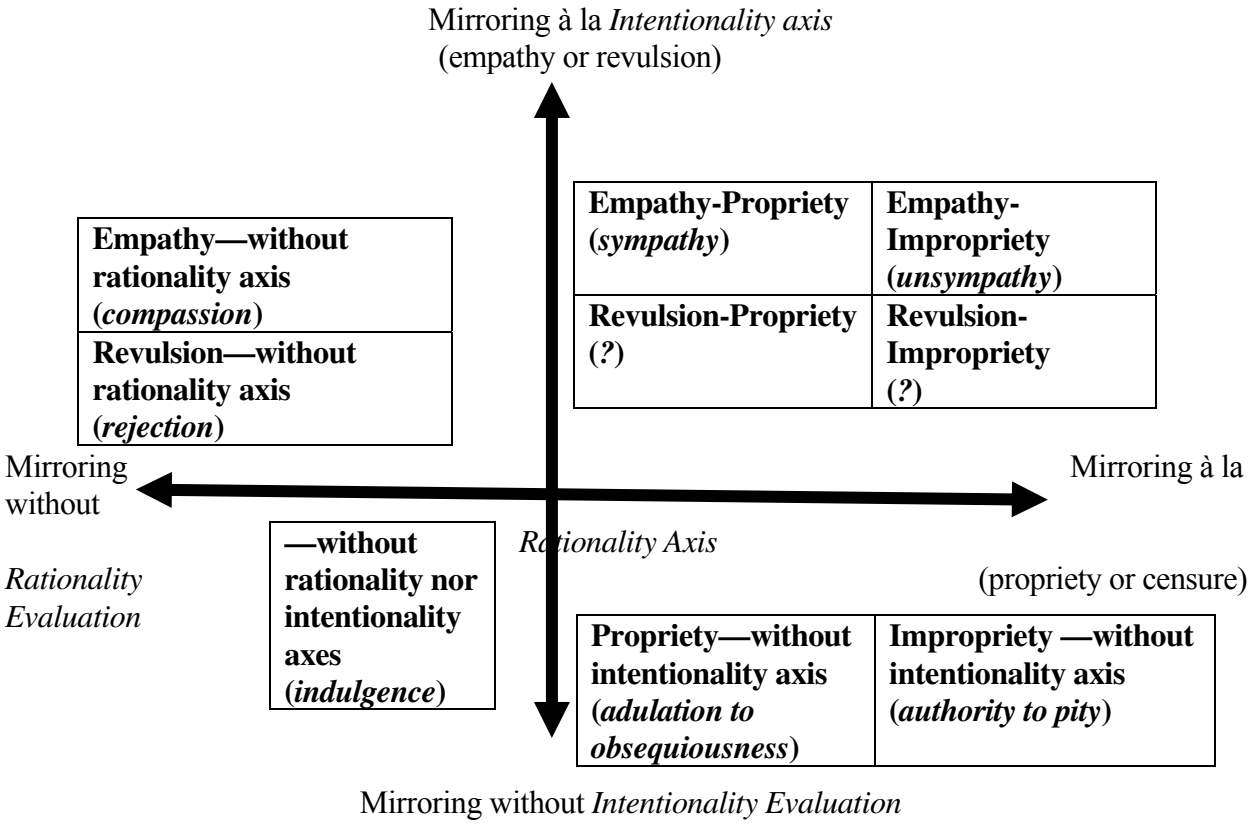

Figure 2: Four Kinds of Fellow-Feeling

quadrant. While "sympathy" is one fellow-feeling from quadrant I, "compassion" from quadrant II. "indulgence" from quadrant III, and "adulation" from quadrant IV.

The emotions of sympathy, compassion, indulgence, and adulation have been studied, among many others, by psychologists [e.g., Ortony *et al.,* 1988; Oatley, 2004; Salovey, 1991; Michael & Haviland, 1993; Portmann, 2000; Ekman, 2003]. Philosophers and others have started to pay serious attention to emotions in light of the collapse of the Cartesian wall separating emotions from rational thinking [Ben-Ze'ev, 2000; Elster, 1999; Damasio, 1994; Ledwig, 2007; Frijda *et al.*, 2000]. Economists have also started to study the emotions, but mainly as strategy in a game setting [e.g., Frank, 1988]. The literature is vast and the terminology is confusing. It cannot be reviewed here.

Nonetheless, Figure 3 shows some of the terms used in the literature insofar as they correspond to the identified four quadrants.[9]

[9] Fontaine [1997] also attempts to clarify the terminological mess. Fontaine contrasts "sympathy," "empathy," and what he called "partial empathy." But these categories are not broad enough to capture what is called here suspension of the intentionality axis, i.e., indulgence and adulation, that arises from imitation. Fontaine's scope is more limited than here because his main focus is on comprehension as defined earlier, or what what Harsanyi and Binmore call "empathy." Fontaine wanted to stay within the economics literature, whose focus is to explain how people understand the constraint budget and utility of each other as they bargain in the market or maximize social welfare function. Fontaine is not interested in ego-centered utility arising from imitation.

Term chosen by this paper	**Equivalent Terms used by others**
"Sympathy"	"Sympathy": Smith [1976] and Scheler [1954] "Fellow Feeling": Smith [1976]; Scheler [1954]
"Indulgence"	"Pity": Nietzsche [2006] "Empathy": Lipps [1960]; Scheler [1954]; Stein [1970]; Heidegger [1962] "Sympathy-as-squeamishness": Sen [1977] "Subjective Preferences": Harsanyi [1977] "Sympathy": Binmore [1994, 1998]
"Compassion"	"Extended Sympathy": Arrow "Empathy": Harsanyi [1977] "Empathy": Binmore [1994, 1998] "Christian Love": Stein [1970] "Mercy": Stein [1970]
"Adulation" ("Pity")	"Imaginative Sympathy": Smith [1976] ("Vanity"; "Pride": Smith [1976])

Figure 3: The Terminological Jumble

6.1 Sympathy

Sympathy is defined as fellow-feeling that arises when the rationality and the intentionality axes are engaged and, further, when the spectator understands as well as approves of the propriety of the principal's action. Given it involves the approbation of propriety, sympathy is *qualified* understanding:

> **Definition:** Sympathy combines empathy and propriety. The spectator, residing in quadrant I, expresses sympathy only when, first, empathizes with the intention of the principal and, second, approves the propriety of the action.

Along the intentionality axis, the spectator empathizes with the principal if the principal is working to enhance wellbeing, whether his utility or the utility of a loved one. Along the rationality axis, the spectator approves the propriety of the action if the principal acts with restraint, i.e., the principal is

not swept away with temptations or myopic emotions.

Even if the spectator cannot empathize with the principal's emotion, the principal's emotion can still be attenuated as a result of the engagement of the rationality axis. The rationality axis is responsible for the mechanism of self-command, not too different from how Smith described it. Such mechanism ensures that the principal, in order to be sure it is acting rationally, must restrain the original emotion. As such, attenuation pathway is guaranteed as long as the rationality axis is engaged, even when the spectator does not empathize with the principal.

If the spectator empathizes with the principal, it does not mean he sympathizes with him or her. While sympathy entails empathy, empathy does not entail sympathy. While the spectator may empathize, the spectator may judge that the principal's action is suboptimal if the principal succumbed to temptation favoring the current self over the interest of a worthy other or over the interest of a future self. In one case, he would be deemed "selfish" while in the other case he would be deemed "imprudent." Such selfishness or imprudence, nonetheless, is motivated by the attempt to improve the welfare, although myopically, of the current self. So, the spectator would be empathic with the selfish principal, but he would be "unsympathetic."

Note, one should not confuse "unsympathy" with the two other possibilities in quadrant I. In these two other possibilities, the spectator finds the intention of the principal revolting if motivated by malice. Nonetheless, the agent can fin the action as either rational or suboptimal. These two cases cannot be explained by Smith's theory. Smith's theory can only account for sympathy or unsympathy as defined here—i.e., only when there is only empathy. Put differently, for Smith, an action is non-understandable only if it is suboptimal, i.e., the agent simply over-reacts or surrenders

to temptation. So, Smith, as stated earlier, conflates sympathy, i.e., approval about propriety, with understanding (empathy), which can involve sympathy or unsympathy. So, for Smith, if the spectator is unable to empathize, it is because the spectator finds the act improper, i.e., finds it suboptimal. Smith did not consider the following: if the spectator is unable to empathize, it is because the spectator finds the act to be revolting, as discussed next.

6.2 Compassion

Compassion is defined as fellow-feeling that arises when the intentionality axis is engaged and, further, the spectator understands (empathizes with) the principal's action. Given it does not involve approbation of propriety, compassion is *pure* understanding:

> **Definition:** Compassion is pure empathy. The spectator, residing in quadrant II, definitely expresses compassion if he empathizes with the principal's intention, i.e., finds understandable because it enhances wellbeing—while withholding judgment as to whether it also efficient (propriety) or suboptimal (impropriety).

Martha Nussbaum defines compassion similarly, i.e., as free from blame which is nothing by the suspension of the rationality axis. The only concern is the reduction of suffering, i.e., improvement of wellbeing:

> The emotion of compassion involves the thought that another creature is suffering significantly, and is not (or not mostly) to blame for that suffering. It does not involve the thought that someone is to blame for that suffering. One may have compassion for the victim of a crime, but one may also have compassion for someone who is dying from disease (in a situation where that vulnerability to disease is nobody's fault). "Humanity" I take to be a similar idea. So compassion omits the essential element of blame for wrongdoing [Nussbaum in Sunstein & Nussbaum, 2004, p. 301].

Does compassion invoke the escalation pathway? This should not be the case, even if the

rationality axis is suspended. When the spectator understands the principal's joy or pain, the spectator is no longer residing in his or her own station. So, the spectator cannot contribute further to the original excitement. While compassion would not attenuate the original excitement, it cannot escalate it. For escalation to take place, the spectator's own experiences and memories must be invoked and add to the original emotion. But then, the spectator's fellow-feeling would not arise from the principal's station, and hence would not be compassion.

On the other hand, the spectator may express revulsion or disgust if he cannot understand or empathize with the action/emotion of the other, even if such emotion involves suffering. For instance, one may not empathize with the suffering of a serial killer, following the fact that one did not empathize with the killer's intention. The revulsion arises for the same reason when one sees someone drinking crankcase oil of his automobile or eating repulsive meant. In all these cases, such actions are revolting because they reduce wellbeing. The reduction of wellbeing can be the wellbeing of others, as in the case of malevolence, or the wellbeing of the self, as in the case cults. If the action is revolting, the spectator would feel "rejection" towards the action—while again ignoring the rationality axis.

When a theoretician conflates the axes of intentionality and rationality is ignored, it is expected for him or her to conflate malevolence (spite) with selfishness. To wit, the sociobiological [e.g., Wilson, 1975] and the economics literature [e.g., Hirshleifer, 1987; Levine, 1998] uses the terms "spite" and "selfishness" more-or-less interchangeably. In light of the TAE hypothesis, we should be able to distinguish them. Spite or malevolence is probably a more complex form of "schadenfreude" (from German) or "epicaricacy" (from Greek). Richard Smith *et al.* [1996; see also

Leach *et al.,* 2002] found that there is a strong link between envy and schadenfreude or, what is known in Australasia, the "tall poppy syndrome" [Feather & Nairn, 2005].[10] Evil is probably the most extreme form of schadenfreude. An evil act is defined as the "joy" experienced by the principal at the sight of the misery of others, when the principal need not have benefited from the act. In contrast, selfishness is an act that the spectator can understand because the intention is to enhance the wellbeing of current self, but when the optimal choice is to take care more of the interest of future self or of the interest of important other. As such, the spectator, or judge within, expresses unsympathy towards selfish actions—while still empathetic with them. This differs from the spectator's expression of rejection towards schadenfreude—where the spectator cannot even understand (i.e., cannot empathize with) the principal's action. Schadenfreude or, its more extreme forms, envy, spite, and malevolence are emotions/acts that the spectator find revolting.

The "joy" that the principal derives in acts of evil differs from the satiation the principal might feel from acts of vengeance. Acts of vengeance are usually instigated by unfairness. In fact, acts of vengeance, as evolutionary game theorists have shown, may enhance wellbeing by helping the group avoid the free-riding problem [see Bowles, 2004, ch. 2; Friedman & Singh, 2003, 2004; Frank, 1988]. With vengeance, agents retaliate even when they do not *ex post* derive any benefit. But such irrational acts, if carried out by a sufficient number of actors, ensures cooperation and, hence, the production of the public good.

To avoid conflating evil with suboptimality, we need to separate the intentionality axis from

[10] Julie Zilko brought the terms "schadenfreude" and "tall poppy syndrome" to my attention. To note, the syndrome is also used to denote the opposite, viz., the quest after egalitarianism and criticism of arrogance and elitism [Peeters, 2004].

the rationality axis. The intentionality axis— which evaluates whether the act is evil—is orthogonal to the rationality axis—which evaluates whether the act is proper. An evil act can be proper (i.e., rational or efficient). And non-evil act can also be improper (i.e., suboptimal). The intentionality and the rationality axes are orthogonal. Like many modern thinkers who conflate the two axes, Smith's conceptual tool kit simply lacks the torch needed to identify evil and how understanding (lack of evil) still may not entail sympathy. The TAE hypothesis allows us to see how rejected acts (evil) can be optimal, while understandable acts (i.e., selfishness) can be suboptimal.

6.3 Indulgence

Indulgence is defined as fellow-feeling that arises when both axes are suspended and, hence, the spectator passes evaluations neither concerning understanding nor approbation of propriety. Given it does not involve approbation of propriety, indulgence is *pure* imitation:

> **Definition:** Indulgence involves neither empathy/revulsion nor propriety/impropriety. The spectator, residing in quadrant III, enjoys the action of the principal *neither* because he or she understands the intention behind it *nor* because he or she finds the action proper. The spectator rather enjoys the action of the principal because it arouse in him his own memories and passions. The spectator does not have to be narcissistic in these enjoyments—although narcissism is usually associated with frequent episodes of indulgence.

Theodor Lipps [1960] defines indulgence similarly, which he incidentally calls "empathy" [see Gladstein, 1987]. The idea of indulgence is how Lipps defines the aesthetic experience. Lipps conceived the aesthetic experience as the projection of one's self-centered emotion on the viewed object:

> Aesthetic enjoyment is a feeling of pleasure of joy in each individual case colored in some specific way and ever different in each new esthetic object—a pleasure caused by viewing the object [Lipps, 1960, p. 374].

So, the aesthetic experience is no different from infectious laughter where one laughs because one hears the laughter of others, i.e., as a result of the "imitation function" of fellow-feeling, without really *understanding* the cause of their laughter.

Other examples of indulgence include the spoiling of a child with toys and gadgets because it allows the parent to enjoy, vicariously, the pleasures of the child. So, the parent would not assess the propriety of the child's enjoyment or the intentionality of the child's action. The parent (spectator) would provide resources to the child (principal) mainly to maximize the parent vicarious utility function.

Indulgence is the only fellow-feeling, out of the four quadrants, that leads to escalation. Given that the rationality axis is suspended, attenuation via self-command is not possible. Further, given that the intentionality axis is suspended, the emotion expressed by the spectator is issued from the spectator's station. As such, the spectator's emotion may have little to do with what the principal actually feels. It has mostly to do with what the principal should feel if the same stimulus or incentive happened to the spectator. So, if the stimulus is a joke told by the principal, the spectator would laugh for his own reasons, which would feed into and stimulate the laughter of the principal. Smith, as noted earlier, discussed why it is always more enjoyable for friends to read a book together or to watch a show together [Ewing *at al.,* 1997].

Indulgence is not limited to enjoyment, though. Indulgence can involve pain, such as squeamishness, which the agent tries to avoid. For instance, one can be squeamish, and not eat meat,

after a visit to the slaughterhouse. One would not eat meat for a week or a month not because one is repulsed, but rather because the thought of blood reminds one of unpleasant experiences.

Further, indulgence can involve escalation of pain which the spectator, amazingly, seeks. The spectator may seek to learn about the suffering of others, not out of compassion (which implies empathy), but rather out of self-centered need to remind the self or re-consume one's own suffering. As the earlier quote from Nietzsche attests, Christianity to him is the key to indulgence (which he calls "suffering") that saps one's ambition and one's will to excel.

As mentioned earlier, theorists as far apart as Hobbes and Becker have relied on indulgence utility to explain altruism. It is proper to call such theories of altruism egocentric. Smith criticized Hobbes's egocentric theory of altruism—a critique that equally applies to Becker's [Khalil, 2001]. Smith argued that sympathy—which also include its corresponding action, altruism—does not stem from ego-centric, "warm glow" pleasures. If it stems from egocentric fellow-feeling, how come, Smith asks, men can have fellow-feeling towards women in labor, when in fact they could never in their own person undergo such an experience. The fact that the man empathizes with the woman as a woman is because the man does not dwell in his self-centered station, but rather he transports himself to the station of the woman:

> A man may sympathize with a woman in child–bed; though it is impossible that he should conceive himself as suffering her pains in his own proper person and character [Smith, 1976, p. 312].

Smith is correct that altruism, once narrowly defined, cannot be based on indulgence. But this does not rule out that in other schemes of income sharing, the motive of the spectator, who shares income with the principal, might be indulgence, i.e., vicarious pleasure rather than

altruism. And such indulgence, given the suspension of the intentionality axis, is usually facilitated by resemblance of traits.[11] Obviously such resemblance of traits, as Smith notes, does not exist between the man and the woman in labor. But resemblance of traits, contrary to Smith, can be the basis of schemes of income sharing other than altruism. Such other schemes include the spoiling of a principal—such as a child—in order for the spectator to indulge himself or herself.

6.4 Adulation

Adulation is defined as fellow-feeling that arises when the intentionality axis is disengaged and, further, the spectator approves of the propriety of the principal's action. Given it involves the approbation of propriety, adulation is *qualified* imitation:

> **Definition:** Adulation is pure propriety. The spectator, residing in quadrant IV, finds the principal's action or achievement worthy of pursuing. Given the suspension of the intentionality axis, the spectator's judgment is not actually about the principal's wellbeing, but rather is about using the principal as an exemplar that can facilitate the spectator's own desire or ambition.

So, similar to indulgence, the spectator does not care about the principal *in his station*. But, in

[11] Trait resemblance plays a critical role in Hume's theory of sympathy. This role complements, as David Levy and Sandra Peart [2004] note, Hume's definition of sympathy as about contagion of emotion or what is called here "indulgence" [Khalil, 2002b]. However, Levy and Peart proceed and advance an interesting, although an indefensible thesis. Namely, Hume's notion of sympathy *necessarily entail* a narrow sense of civil society than encompasses only the Europeans; while Smith's notion of sympathy, i.e., grounded on humanity, has a broad notion of civil society that encompasses all mankind. It is correct that Hume's civil society is narrower than Smith's. But this difference cannot be traced to the difference in their definition of sympathy. The issue of trait resemblance, important for Hume, need not, *a priori,* be restricted to Europeans—even though Hume does restrict it to Europeans. Humans, and even primates, have bounded by trait resemblance. The boundary of civil society is an issue that cannot be

dissimilar to indulgence, the spectator uses the principal in order to live his own live vicariously, i.e., as a vehicle for him to judge and boost his own self-regarding or self-worth.

As discussed earlier, when the principal acts properly, another source of satisfaction arises besides the maximized utility. Namely, the principal experiences self-regarding or self-worth for taking the optimum action which is aside from the utility derived from such action. Such self-worth is a reflection of one's accomplishment or success. The principal's joy of self-worth seems to be a more potent element in the spectator's vicarious enjoyment than the principal's joy of the maximized utility. That is, the second source of satisfaction, self-worth, seems to play a more prominent role in quadrant IV than in quadrant I, when both quadrants are engaging the same rationality axis. For one thing, in quadrant IV, the spectator is judging matters from his self-centered station, i.e., when the intentionality axis is suspended, unlike quadrant I when the intentionality axis is engaged. As such, the spectator wants to evaluate his own self-worth, i.e., experience the secondary emotion resulting from doing he proper action. So, he enjoys his self-worth vicariously by imagining the self-worth of the admired principal is happening to his own satin. This situation may occasion jealousy rather than admiration. But we do not need to enter into such details here [see Khalil, 1996]. The basic point here is that the engaged rationality axis in quadrant IV does not perform exactly the same function as it does in quadrant I, and for a good reason. Given that the intentionality axis in quadrant IV is suspended, i.e., the spectator is self-centered, the engagement of the rationality axis amounts to emphasizing the second source of satisfaction, i.e., the sense of self-worth that arises from taking effective (rational) decisions.

---

resolved by the study of fellow feeling. The two are independent questions [Khalil, 2007b].

But does adulation, similar to indulgence, lead to the escalation of the principal's sense of worth, i.e., Hume's "accelerator pedal" of fellow-feeling? Adulation usually does not involve escalation. The excitement of the spectator who adores the principal may instigate the principal to hold back. In order for the principal to admired and adored, the principal must exercise propriety. Otherwise, much of the mystique that is so adored would vanish. This is the case because the principal is fully aware that the spectator is not really interested in knowing about the intention of the principal. The spectator is rather self-centered.

The spectator is evaluating his or her own standing in relation to the standing of coveted positions of others. Such positions are the externalized reference point for what one believes to be his ability and his desire. The beliefs about the self and what it can desire are basically about what James Heckman calls "non-cognitive beliefs" [Heckman, 2007, in Heckman & Krueger, 2003]. While cognitive beliefs are about one's environment, noncognitive beliefs concern one's ability which explains desire, ambition, and internal motivation [Khalil, 2007c,d].

If one's quest or desire concerns etiquettes, the spectator asks whether the way the principal walks, eats, dresses, and so on, is more elegant than the way he or she walks, eats, dresses, and so on. The evaluation of one's etiquettes is not trivial as supposed at first look. It indicates one's care about health, risk, and so on. One's quest or desire can be wealth, knowledge, beauty, sociability, and so on. Whatever is the metric, the spectator measures his accomplishments in relation to the principal's or, what is the same thing, in relation to his own goal. Both yardsticks are the same. The spectator, after all, selects the principal, or the social reference group, against which he or she would like to gauge his own performance.

If the spectator finds that the principal, with regard to the selected metric, has a higher achievement as a result of prudence and tenacious effort, the spectator would use the principal as an exemplar in order to exercise similar prudence and tenacity and achieve a similar standing. Such judgment of standing or status is more involved. Factors such as luck and natural aptitudes play a role [Khalil, 1996], which we will ignore here for simplicity. We will focus only on how the spectator judges relative standing, as if accomplishment is purely the outcome of prudence and tenacity. Our spectator may experience jealousy towards the principal as the spectator tries hard to attain his own desire that he sees so perfectly achieved by the object of his jealousy, the principal. The jealousy, though, is usually mixed with adulation especially when the jealous spectator starts to believe that he cannot attain what he truly desires.

On the other hand, if the spectator finds that the principal has a lower standing, i.e., the principal failed to act prudently and tenaciously, the spectator would feel pre-eminent or has a lead over the principal. Such feeling, following Adam Smith [1976, p. 50; Khalil, 2002a; 2005] is called here "authority." The spectator judges himself as superior vis-à-vis the principal with regard to the selected metric. The term "authority" is not used here in a pejorative sense, i.e., as if the spectator is patronizing or treating the principal in a condescending manner. When authority becomes arrogance or patronizing is called "pity." While pity presupposes authority, it involves another element discussed below, viz., "elitism" in the sense of arrogance and snobbery. The term "authority" as used here denote simply the fact that a mentor—such as a teacher or a parent—can act as an exemplar with respect to propriety. Such mentoring may involve influencing the development of the principal's utility function. So the mentoring is not simply about increasing efficiency in the sense

of providing either information or precommitments to assist the principal with self-command (prudence) in the face of temptations. So authority or mentoring differs from what Cass Sunstein and Richard Thaler's [2003] call "libertarian paternalism." Such paternalism is simply about enhancing efficiency by providing either better information, e.g., what food is good for health, or precommitments, such as "forced" saving schemes.

So, the spectator in quadrant IV can express, towards the principal, the fellow-feeling of jealousy that may lead to adulation (in case the principal acts with propriety and tenacity) or authority (in case the principal is lacking in achievement). The adulation/authority twin perform same function: the assessment of one's self or others as one pursues the object of desire.

René Girard [1972] regarded, reminiscent of Nietzsche, *desire* as the defining question of the human condition. Broadly speaking, except for the theories of Nietzsche and a few others, modern social theory has neglected the role of desire as the entry point of theorizing about the human condition. Of course, most theories eventually discuss desire. But the point is whether desire acts as the organizing principal to make sense of diverse phenomena. Modern social theory is mainly concerned with the social contract in light of competing interests and the problem of free-riding. But humans still have to deal with desire even if they live as Robinson Crusoe. Girard's work show how the frustration of desire leads mortals to make Gods of each other. And in this act of adulation/authority, the lower status agent do not want to know that the emperor or the Gods have no clothes, as much as the people acting as authority do not want reveal themselves naked.

Karl Marx [1973] discussed at length adulation/authority relation that binds the chieftains, kings and emperors with their subjects. Marx, though, restricted such adulation to pre-capitalist

social formation. Marx argued that the root of such adulation, which I called "rank fetishism," is the fear of nature [se Khalil, 1992]. Marx was typical of modern social theory. He thought that adulation would whither away with the rise of capitalist mode of production because of technological progress, what he called the advancement of "forces of production." The advancement allows humans to control nature. Consequently, humans would no longer be scared with the rise of forces of production (technology). Thus, they would no longer tend to make Gods of mortals.

Smith was not a modernist as Marx. He did not think that concern with rank and status would vanish with the rise commercial society. In his analysis of the origin of rank, Smith [1976, pp. 50-62] rather anticipates Nietzsche and Girard. This is not the place to elaborate his theory of authority, which challenges directly social contract theory of the socialist tradition of Jean-Jacques Rousseau as well as of classical liberalism stretching from Thomas Hobbes, John Lock, to James Buchanan [see Khalil, 1998; 2002a, 2005]. Stated briefly, for Smith, humans, *all* humans, would rather choose death over leading a life empty of desire, i.e., the ambition to reach higher ranking goals. On the other hand, Smith was a great admirer of stoic philosophy that emphasized that nature has no meaning and that ambition/desire is the road to misery. Along the legacy of the stoics, Smith advocated the virtue of "self-love" which is about content, i.e., living a desire-free life. There is a debate about the extent to which Smith was a stoic [see Clarke, 2000]. Nonetheless, he put forward the observation, as a scientist, that humans cannot be stoic, i.e., they cannot lead a life that is desire-free.

For Smith, while people are driven by desire, most of them realize that they cannot attain their desire. So, they adulate other humans that seem to them more successful than themselves.

Such adulation, the lower-rank spectators are not really sympathetic with the welfare of the rich and famous—because they are not engaging the intentionality axis. The spectators are rather operating from their own, self-centered fellow-feeling. So, the news about the more successful agents, i.e., the ones judged to embody desired goals according to the rationality axis, become the object of vicarious enjoyment.

Smith seems to be aware that such vicarious enjoyment, i.e., adulation, is different from his concept of sympathy. This is the case because he called adulation "imaginative sympathy." But Smith never tried to connect his concept of sympathy with adulation. In an earlier analysis, I called adulation in Smith "vicarious sympathy" [Khalil, 2002a, 2005]. I thought the term "vicarious sympathy" is better indicative of adulation than Smith's "imaginative sympathy. However, the term "vicarious sympathy" is, in light of the TEA hypothesis, is inadequate if we want to distinguish "adulation" from "indulgence"—since indulgence (quadrant III) also involves vicarious enjoyment.

To illustrate the difference between adulation and sympathy, let us examine the enormous "sympathy" accompanying the imprisonment of Paris Hilton. Ms. Hilton, a 26-year old heiress of the Hilton hotel fortune, is famous for being famous. So her achievement is not actually examined on her own station; they are rather the fancy of spectators of quadrant IV. The picture of Ms. Hilton splashed the front pages of newspapers around the world as she arrived, in early June 2007, at the Century Regional Detention Facility in suburban Los Angeles to serve a 45-day prison sentence for violating probation in an alcohol-related reckless driving case. Why all this interest and commotion for a 26-year old woman going to jail, in which she is expected to serve 23 days? Why even a website was set up on her behalf by fans to start a petition asking the Governor of California to

pardon her? Is it sympathy, which would invoke the evaluation of her intentions? As painful as the jail ordeal would be for her, there are more horrific ordeals that women undergo everyday in Southeast Asia with the slave-sex trade, and the more agonizing ordeals that women undergo in many poor African countries, viz., they have to take care of family members who have AIDS while they themselves are also infected with AIDS. If the fans of Ms. Hilton, and the wider public, are motivated by sympathy, they would have instead spent their resources on the problems of sex-slave trade and AIDS.[12]

Likewise, adult men have cried when Princess Diane was killed in automobile crash—when they did not even cry or feel the same intensity of loss when they their parents passed away. The fascination with celebrity cannot stem from sympathy, as Smith long ago noted. It must be related to frustrated desire, where there is a judgment of what one can desire. Such a judgment leads to the ranking of people, where the higher rank is worshiped and venerated. Marx was wrong. The advancement of capitalist production failed to free us from rank fetishism and status inequality as Marx predicted. (In contrast, Marx predicted that the advancement of socialism would free us from *income* inequality).

Status inequality arouses the lower rank person to adulate the higher rank. The higher rank,

---

[12] As "Times online" (May 8, 2007) states:

> Jail-bound socialite Paris Hilton urged fans today to sign a petition to pardon her "mistake"…
> The petition reads: "Paris Whitney Hilton is an American celebrity and socialite. She is an heiress to a share of the Hilton Hotel fortune, as well as to the real estate fortune of her father Richard Hilton. She provides hope for young people all over the U.S. and the world. She provides beauty and excitement to (most of) our otherwise mundane lives."

if generous, usually reciprocates with authority as discussed above. Authority need not entail condescension, patronizing behavior, arrogance, or in short pity.

But in many cases authority may lead to pity fellow-feeling. In this case, lower status people are not only expected to adulate the higher status ones, but they are also supposed to venerate them to the point of obsequiousness. When higher status agents express pity, lower status agent should express what is called here "obsequiousness." The twin fellow-feeling of *pity/obsequiousness* is not the product of simple status inequality. Simple status inequality generates the twin fellow-feeling of *authority/adulation*.

It is conjectured here that the pity/obsequiousness twin is rather the product of status inequality mixed with another institution, namely, elitism. There is no consensus on how to define elitism. But it is more than simply Plato's idea that the people with the greatest expertise, i.e., authority, should be given greater role in governing. As defined here, elitism is the institution, belief, or ideology that people could never, even if they try, become equal. It might be true that people can never, even if they try, be equal with regard to one particular talent—such as music or mathematics—as opposed to another—such a poetry or bicycling. But elitism is more than such an innocuous hypothesis. Elitism involves rather a judgment about the "whole" individual. So, lower rank individuals are inherently lower than higher rank individuals "on the whole," i.e., with regard to all or almost all talents. Thus, lower rank individuals are almost confined to their station in life, irrespective of their effort. This is the case because there is, supposedly, a deep factor, traced usually to culture and sometimes to biology, which ranks individuals "on the whole." The terms "snobbery"

and "arrogance" are other, although pejorative, names of elitism.

It is outside our concern here whether psychological, anthropological, and biological findings support the ideology of elitism. What concerns us here is that once status inequality is mixed with elitist ideology, the result is usually the caste system, racial segregation, or stiff social segregation based on other kind of group identity such as religion, ethnicity, accent, and so on. To wit, we can have "pure elitism"—where the snobbery is not even corroborated with actual differences in ability [see Khalil, 2007a]. Racial segregation might express such pure elitism. The racial segregation setting of the novel *To Kill a Mockingbird*, mentioned at the outset, between blacks and even poorer whites might be such institution of pure elitism. Tom, given the institutional matrix of racial segregation, should not only adulate white people. He should also bow to them obsequiously. In return, white people would express not only authority, but also pity. So, pity can be defined as the fellow-feeling of authority that is mixed with patronization or elitism or it can be the product of pure elitism.

While Smith used the term "pity," he used it interchangeably with "compassion" to denote general fellow-feeling. Nonetheless, Smith did not miss an opportunity to describe and criticize ostentatious and arrogant behavior, which is responsible for pity, and its twin fellow-feeling, obsequiousness. To wit, the terms "ostentatious" and "obsequious" are often encountered in *The Theory of Moral Sentiments*. To note, Smith did not delineate between "authority" and its mutilated form, "pity" as conjectured here—i.e., delineate between status inequality, on one hand, and status inequality mixed with arrogance or elitism, which is responsible for social segregation, on the other. For Smith [1976, pp. 255-259; see Khalil, 1996, 2000a], arrogance is found in people who are

inflicted with "weakness of character," i.e., people who are anxious about their standing in the pecking order of society. Smith skillfully distinguished between two "flavors" of arrogant, weak men: the "vain man" and the "proud man." Both exhibit self-aggrandizement. While the vain man is too ready to display his accomplishments in order to remind lower-ranking agents that they cannot reach his rank; the proud man is even too proud even to bother to display his accomplishments.

In short, status inequality, which engenders the adulation/authority twin, can take place without elitism, and *vice versa*. But with elitism, we have stiff social segregation such as the caste system. Such system can help breed inequalities which escalates the system. Nonetheless, stiff institutions of social segregation differ from racial hatred as expressed in the Holocaust. The distinction between racial segregation and racial hatred cannot be demarcated without the TAE hypothesis.

Further, status inequality need not be thrown away along with elitism in order to undermine the obsequiousness/pity twin. To throw away status inequality with elitism would amount to throwing the baby away with the bathwater. Humans may never be able to avoid status inequality as assumed by Marx. On the other hand, vibrant societies can be free of elitism. In any case, we need the concept of status inequality and, correspondingly, the adulation/authority fellow-feeling if we ever want to make sense of *assabiya*—the Arabic term that Ibn Khaldûn [1967] uses to denote tribalism, group identity, allegiance, team spirit, or, what is currently called, nationalism.

7. How far is *Assabiya* apart from other "Social Preferences"?

It is conjectured here that the adulation/authority twin might be the elementary building block for the study of allegiance, group solidarity, or the bond that unites the citizens of the state. In fact, Adam Smith [1976. pp. 50-62] argued that status inequality, i.e., the adulation/authority twin, is the corner stone of understanding political authority.

Smith [1978] spent a great deal of effort in analyzing the nature of political authority. Smith directly criticized John Locke's theory [Khalil, 1998]. The social contract idea, based on interests, simply misses the role of desire and, hence, fails to grasp the nature of authority. For Smith [1976, p. 50] desire is the entry point of analysis if we want to explain authority. What matters for Smith's analysis is that desire is often frustrated. Frustrated desire is nonetheless fulfilled though adulation, as discussed above. Adulation amounts to the fusion of egos, where the spectator identifies his ego with the imagined ego of the team, producing what is usually called "team spirit." But team spirit may not be different from how a sports fan identifies with a sports team or a movie viewer identifies with the hero of a film. So, the fusion of egos cannot be the whole story of political allegiance. To wit, Smith argues that there is another element, aside from authority, that is needed in order to explain political allegiance.

Stated briefly, the adulation/authority twin must be combined with the principle of interest or utility [see Khalil, 2002a, 2005]. Once authority is combined with interest, the adulation/authority twin is transformed into allegiance, group solidarity, or, in short, a*ssabiya*. Such *assabiya* prompts spectators to cry when they see their king, touch their flag, or hear the national anthem.[13]

[13] In fact, one major aspect of Smith's notion of the invisible hand in *The Theory of Moral Sentiments* [Khalil, 2000b] is about the spontaneous rise of political order. He discusses a length how the myopic sentiments of adulation/authority gives rise, once combined with interest, to allegiance. Such allegiance affords political order which would whither away if there is no *assabiya*.

How can we explain such nationalist or *assabiya* emotion? Is it the same as altruism and fairness, which also benefit other group members? Actually, motives such as *assabiya* (under the guise of group identity), altruism, and fairness have been receiving great attention in the literature under terms such as "social preferences" and "prosocial preferences" [e.g., Gintis, 2003; Bowles, 2004; Gintis *et al.*, 2005; Bénabou & Tirole, 2006]. One has to be careful, though, not to suggest that altruism and fairness have the same standing as *assabiya*, and just lump them all as "social preferences." In light of the TAE hypothesis, we should not use the same model to conceive *assabiya* as the once used to conceive altruism and fairness.

Stated briefly, altruism and fairness are ultimately about the evaluation of action in light of the intentionality axis. That is, the agent is trying, in both altruism and fairness, to enhance wellbeing as evaluated by the intentionality axis. In contrast, with the adulation/authority twin, responsible for *assabiya*, the intention of the principal is not under consideration to start with. It does not matter whether the principal's action is wellbeing or not. What matters is how a lower status spectator feels by imagining the accomplishments of the great and powerful as it they are his or her own. So, when a spectator acts according to allegiance, the action is not the same as when he or she acts altruistically or fairly.

9. Conclusion

Hume's fellow-feeling paradox is solved. The same primitive fellow-feeling can attenuate the original emotion, but only if the rationality axis is engaged. It also can escalate the original emotion,

as in the case of the death of Mr. Morales, but only if the rationality and the intentionality axes are disengaged. This solution cannot be possible unless we treat the two axes as orthogonal—the core thesis of the TAE hypothesis.

Further, the TAE hypothesis allows us to solve the mirror-neuron paradox by identifying the "understanding function" and the "imitation function" of fellow feeling. Both are mirror images along the intentionality axis. The "understanding function" is possible only when the intentionality axis is engaged. In this instance, fellow-feeling can take the forms, among other things, of compassion and sympathy. The "imitation function" is possible only when the intentionality axis is disengaged. In this instance, fellow-feeling can take the forms, among other things, of indulgence and adulation. Adulation towards higher rank can develop into obsequiousness, and in return the higher rank can express pity. Without the TAE hypothesis, it would not be possible to distinguish sympathy from pity, which was partially responsible for the wrongful conviction of Tom Robinson in *To Kill a Mockingbird*.

This paper showed how comprehension of the acts of serial killers does not entail understanding (empathy). In turn, the understanding, e.g., of war waging does not entail sympathy with the attacker. These three layers—comprehension, understanding, and sympathy—cannot be distinguished without the demarcation between the intentionality axis and the rationality axis, the core of the TAE hypothesis.

Furthermore, selfishness differs from spite (evil) which, in turn, differs from snobbery. A principal acts selfishly when he or she enhances wellbeing—but it is still suboptimal given his or her regard for the wellbeing of loved ones including future self. A principal acts out of spite or malice

when he or she hurts someone without enhancing wellbeing. A principal acts with snobbery when he assumes, first, whether rightly or wrongly, that he has a superior ability and, second, others cannot reach his status even if they try. These three anti-social sentiments—selfishness, spite, and snobbery—cannot be distinguished without the demarcation between the intentionality axis and the rationality axis, the central proposal of this paper.

Economists have exclusively focused on the rationality axis—totally ignoring the intentionality axis. So, the tools of economics cannot explore the problem of evil and how it differs from snobbery. Even with the rise of behavioral economics, there is little hope that it will be able to tackle the question of evil. Behavioral economists are challenging, among other things, the revealed preference axiom. But even if one disputes the axiom, and admits that agents do not behave rationally, this does not invoke the intentionality axis.

Economics is not alone in ignoring the intentionality axis. It is actually the mark of the rise of modern social science—on the shoulders of Machiavelli, Hobbes, Locke, Rousseau, and Marx—to be antagonistic to the question of evil. Modern social science views the "economic problem" or the "human condition" as about the engineering of the best institutions that concern issues raised exclusively by the rationality axis, i.e., the benefits of competition and cooperation among rival or complementary interests. So, if one person hurts another it is only because the perpetrator is pursuing, efficiently or not efficiently, his or her wellbeing. So, we have atoms that collide simply as a result of the pursuit of wellbeing. The modern palace of social science has no room, even a closet, for the study of evil. It is hoped that this essay has opened a window in the palace that is wide enough to entice further scientific study of evil.

References

Arrow, Kenneth J. "Extended Sympathy and the Possibility of Social Choice." *American Economic Review*, Papers and Proceedings, February 1977, 67:1, pp. 219-225.

Baron-Cohen, S. *Mindblindness: An Essay on Autism and Theory of Mind*. Cambride, MA: MIT Press, 1995.

_____, H. Tager-Flusberg, and D.J. Cohen (eds.) *Understanding Other Minds: Perspectives from Developmental Cognitive Neuroscience*, 2nd ed. Oxford: Oxford University Press, 2000.

Basch, Michael. "Empathic Understanding: A Review of the Concept and Some Theoretical Considerations." *Journal of the American Psychoanalytic Association,* 1983, 31:1, pp. 101-26.

Becker, Gary S. "Altruism in the Family and Selfishness in the Market Place." *Economica*, February 1981, 48, pp. 1-15.

_____. "A Note on Restaurant Pricing and Other Examples of Social Influences on Price**."** *Journal of Political Economy*, October 1991, 99:5, pp. 1109-1116.

_____. *Accounting for Tastes*. Cambridge, MA: Harvard University Press, 1996.

_____ and Kevin M. Murphy. "A Simple Theory of Advertising as a Good or Bad." Quarterly Journal of Economics, November 1993, 108:4, pp. 941-964.

Bénabou, Roland and Jean Tirole. "Incentives and Prosocial Behavior." *American Economic Review*, December 2006, 96:5, pp. 1652-1678.

Ben-Ze'ev, Aaron. *The Subtlety of Emotions*. Cambridge, MA: MIT Press, 2000.

Bernheim, B. Douglas and Oded Stark. "Altruism Within the Family Reconsidered: Do Nice Guys Finish Last?" *American Economic Review*, 1988, 78: 1034–1045.

Binmore, Ken. *Game Theory and the Social Contract. Volume 1: Playing Fair*. MIT Press, 1994.

_____. *Game Theory and the Social Contract. Volume 2: Just Playing*. MIT Press, 1998.

Bowles, Samuel. *Microeconomics: Behavior, Institutions and Evolution*. New York: Russell Sage; Princeton, N.J.: Princeton University Press, 2004.

Buccino, Giovanni, Fausta Lui , Nicola Canessa, Ilaria Patteri, Giovanna Lagravinese, Francesca Benuzzi, Carlo A. Porro, Giacomo Rizzolatti. "Neural Circuits Involved in the Recognition of

Actions Performed by Nonconspecifics: An fMRI Study." *Journal of Cognitive Neuroscience*, 2004; 16, pp. 114-126.

Burt, Austin and Robert Trivers. *Genes in Conflict: The Biology of Selfish Genetic Elements*. Cambridge, MA: Harvard University Press, 2006.

Clarke, Peter H. "Adam Smith, Stoicism and Religion in the Eighteenth Century." *History of the Human Sciences*, 2000, 13:4, pp. 49–72.

Damasio, Antonio R. *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: G.P. Putnam, 1994.

Darwall, Stephen. "Empathy, Sympathy, Care." In Darwall's *Welfare and Rational Care*. Princeton, NJ: Princeton University Press, 2002, ch. 3.

_____. *The Second-Person Standpoint: Morality, Respect, and Accountability*. Cambridge, MA: Harvard University Press, 2006.

Dupuy, Jean-Pierre. "Intersubjectivity and Embodiment." *Journal of Bioeconomics*, 2004, 6:3, pp. 275-294.

_____ "Invidious Sympathy in *The Theory of Moral Sentiment*." *Adam Smith Review*, 2006, 2, pp. 98-123.

Ekman, Paul. *Emotions Revealed: Understanding Faces and Feelings*. London: Weidenfeld & Nicolson, 2003.

Elster, Jon. *Alchemies of the Mind: Rationality and the Emotions*. Cambridge: Cambridge University Press, 1999.

Ewing, Michael T., Colin Jevons, and Elias Khalil. **"**Brand Senescence: Towards a Developmental Theory of Brand Mortality." A working paper, 2007.

Feather, N.T. and Nairn, Katherine. "Resentment, Envy, Schadenfreude, and Sympathy: Effects of own and other's Deserved and Undeserved Status." *Australian Journal of Psychology*, August 2005, 57:2, pp. 87-102.

Fontaine, Phillipe. "Recognition and Economic Behavior: Sympathy and Empathy in Historical Perspective." *Economics and Philosophy*, 1997, 13, pp. 261-280.

_____. "The Changing Place of Empathy in Welfare Economics." *History of Political Economy*, Fall 2001, 33:3, pp. 387-409.

Frank, Robert H. *Passions Within Reason: The Strategic Role of the Emotions*. New York: W.W. Norton, 1988.

_____. *Microeconomics and Behavior*, 6th ed. New York: McGraw Hill Irwin 2006.

Friedman, Daniel and Nirvikar Singh. "Negative Reciprocity: The Coevolution of Memes and Genes." Department of Economics working papers #560, University of California at Santa Cruz, 2003.

_____. "Vengefulness Evolves in Small Groups." Department of Economics working papers #559, University of California at Santa Cruz, 2004.

Frijda, Nico H., Antony S.R. Manstead, Sacha Bem (eds.). *Emotions and Beliefs: How Feelings Influence Thoughts*. Paris: Editions de la Maison des Sciences de l'Homme; Cambridge: Cambridge University Press, 2000.

Fudenberg, Drew and David K. Levine. "A Dual-Self Model of Impulse Control." *American Economic Review*, December 2006, 96:5, pp. 1449-1476.

Gallese, V., C. Keysers, and G. Rizzolatti. "A Unifying View of the Basis of Social Cognition." *Trends in Cognitive Sciences*, 2004, 8:9, pp. 396-403.

Gintis, Herbert. "Solving the Puzzle of Prosociality." *Rationality and Society*, 2003, 15, pp. 155-187.

_____, Samuel Bowles, Robert Boyd, and Ernst Fehr (eds.). *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life.* Cambridge, MA: MIT Press, 2005.

Girard, René. *Deceit, Desire and the Novel: Self and Other in Literary Structure*, translated by Yvonne Freccero. Baltimore: The Johns Hopkins University Press, 1976.

Gladstein, Gerald A. "The Historical Roots of Contemporary Empathy Research." *Journal of the History of the Behavioral Sciences*, 1984, 20:1, pp. 38-59.

_____ *et al. Empathy and Counseling: Explorations in Theory and Research*. New York: Springer-Verlag, 1987.

Gordon, Robert. "Sympathy, Simulation, and the Impartial Spectator." *Ethics*, Summer 1995. (Reprinted in Larry May, Marilyn Friedman, and Andy Clark *(eds.) Mind and Morals: Essays on Ethics and Cognitive Science* Cambridge, MA: MIT Press, 1996.)

Haig, David. "Genetic Conflicts in Human Pregnancy." *Quarterly Review of Biology*, December 1993, 68:4, pp. 495-532.

_____. "On Intrapersonal Reciprocity." *Evolution and Human Behavior*, 2004, 24, pp. 418-425.

Harsanyi, John. "Cardinal Welfare, Individualistic Ethics and Interpersonal Comparisons of Utility." *Journal of Political Economy*, 1955, 63, pp. 309–321.

_____. *Rational Behavior and Bargaining Equilibrium in Games and Social Situations.* Cambridge: Cambridge University Press, 1977.

Heckman, James J. "The Economics, Technology and Neuroscience of Human Capability Formation." Institute for the Study of Labor (IZA), Bonn, Discussion Paper No. 2875, June 2007.

_____ and Alan B. Krueger. *Inequality in America: What Role for Human Capital Policies?* The Alvin Hansen Symposium on Public Policy, Harvard University; edited with an introduction by Benjamin M. Friedman. Cambridge, MA: MIT Press, 2003.

Heidegger, Martin. *Being and Time,* trans. John Macquarrie, Edward Robinson. New York: Harper, 1962.

Hirshleifer, Jack. "On the Emotions as Guarantors of Threats and Promises." In John Dupré (ed.) *The Latest on the Best: Essays on Evolution and Optimality*. Cambridge, MA: MIT Press, 1987, pp. 307-326.

Hollis, Martin. *Trust Within Reason.* Cambridge University Press, 1998.

Hurley, Susan and Nick Chater (eds.). *Perspectives on Imitation: From Neuroscience to Social Science*, 2 vols. Cambridge, MA: MIT Press, 2005.

Ibn Khaldûn. *The Muqaddimah: An Introduction to History*, 3 vols., 2nd edition. Trans. From the Arabic by Franz Rosenthal, Bollingen Series XLIII. Princeton: Princeton University Press, 1967.

Karni, Edi, and David Schmeidler. "Fixed Preferences and Changing Tastes." *American Economic Review, Papers and Proceedings*, May 1990, 80:2, pp. 262-267.

Khalil, Elias L. "Beyond Self-Interest and Altruism: A Reconstruction of Adam Smith's Theory of Human Conduct." *Economics and Philosophy*, October 1990, 6:2, pp. 255-273.

_____. "Nature and Abstract Labor in Marx." *Social Concept*, June 1992, 6:2, pp. 91-117.

_____. "Respect, Admiration, Aggrandizement: Adam Smith as Economic Psychologist." *Journal of Economic Psychology*, September 1996, 17:5, pp. 555-577.

_____. "Is Justice the Primary Feature of the State? Adam Smith's Critique of Social Contract Theory." *European Journal of Law and Economics*, November 1998, 6:3, pp. 215-230.

_____. "Symbolic Products: Prestige, Pride and Identity Goods." *Theory and Decision*, August 2000a, 49:1, pp. 53-77.

_____. "Beyond Natural Selection and Divine Intervention: The Lamarckian Implication of Adam Smith's Invisible Hand." *Journal of Evolutionary Economics*, 2000b, 10:4, pp. 373-393.

_____. "Adam Smith and Three Theories of Altruism." *Recherches* Économiques *de Louvain – Louvain Economic Review*, 2001, 67:4, pp. 421-435.

_____. "Is Adam Smith Liberal?" *Journal of Institutional and Theoretical Economics*, December 2002a, 158:4, pp. 664-694.

_____. "Similarity vs. Familiarity: When Empathy Becomes Egocentric." *Behavioral and Brain Sciences,* February 2002b, 25:1, p. 41.

_____. "What is Altruism?" *Journal of Economic Psychology*, February 2004, 25:1, pp. 97-123.

_____. "An Anatomy of Authority: Adam Smith as Political Theorist." *Cambridge Journal of Economics*, January 2005, 29:1, pp. 57-71.

_____. "Introduction: Smith the Hedgehog." *Adam Smith Review*, 2006, 2, pp. 3-20.

_____. "Rank Fetishism and Corruption: Why the Adulation of Vacuous Celebrities can Enhance Productivity." A working paper, 2007a.

_____. "The Moral Justification of Enslavement." A working paper, 2007b.

_____. "The Cognitivist Fallacy: The Heckman Puzzle." A working paper, 2007c.

_____. "Making Sense of Behavioral Anomalies: Four Kinds of Presentational Effects." A working paper, 2007d.

_____. "The Fellow-Feeling Paradox: Hume, Smith, and the Commitment Problem." A working paper, 2007e.

_____. "Why Shame Cannot be a Price." A working paper, 2007f.

_____. "Tastes." *International Encyclopedia of the Social Sciences,* 2nd ed., Macmillan Reference USA, Editor-in-Chief: William A. Darity, 2008, in press.

Leach, C.W., R. Spears, N. Branscombe, and B. Doosje. "Malicious Pleasure: Schadenfreude at the Suffering of Another Group." *Journal of Personality and Social Psychology*, 2002, 84, pp. 932-943.

Ledwig, Marion. *Emotions: Their Rationality & Consistency*. New York: Peter Lang, 2007.

Lee, Harper. *To Kill a Mockingbird*. London: Mandarin, (1960) 1989.

Levine, David K. "Modeling Altruism and Spitefulness in Experiments." *Review of Economic Dynamics*, 1998, 1, pp. 593-622.

Levy, David M. *How the Dismal Science Got its Name: Classical Economics & the Ur-text of Racial Politics.* Ann Arbor: University of Michigan Press, 2001.

_____ and Sandra J. Peart. "Sympathy and Approbation in Hume and Smith: A Solution to the other Rational Species Problem." *Economics and Philosophy*, October 2004, 20:2, pp. 331-349.

Lewis, Michael and Jeannette M. Haviland (eds.). *Handbook of Emotions*. New York: Guilford Press, 1993.

Lipps, Theodor. "Empathy, Inner Imitation, and Sense-Feelings." In Melvin Rader (ed.) *A Modern Book of Esthetics: An Anthology*, 3rd ed. New York: Holt, Rinehart and Winston, (1935) 1960, pp. 374-382.

Marx, Karl. *Grundrisse*, translated with a foreword by Martin Nicolaus. New York: Vintage, 1973.

_____. *Capital,* vol. 1, intro. by Ernest Mandel. Harmondsworth: Penguin and London: New Left Books, 1976.

McCloskey, Deirdre N. *The Bourgeois Virtues: Ethics for an Age of Commerce*. Chicago: University of Chicago Press, 2006.

Meltzoff, A.N. and A. Gopnik. "The Role of Imitation in Understanding Persons and Developing a Theory of Mind." In S. Baron-Cohen, H. Tager-Flusberg, and D.J. Cohen (eds.) *Understanding Other Minds: Perspectives from Autism.* Oxford: Oxford University Press, 1993.

Miller, William Ian. *The Anatomy of Disgust*. Cambridge, MA: Harvard University Press, 1997.

Nietzsche, Friedrich. *The Anti-Christ: Curse on Christianity*. In *The Nietzsche Reader*, ed. By Keith Ansell Pearson and Duncan Large. Malden, MA: Blackwell, 2006.

Oatley, Keith. *Emotions: A Brief History*. Malden, MA: Blackwell, 2004

Ortony, A., G.L. Clore, and A. Collins. *Cognitive Structure of Emotions*. New York: Cambridge University Press, 1988.

Peeters, Bert. "Tall Poppies and Egalitarianism in Australian Discourse: From Key Word to Cultural Value." *English World-Wide*, 2004, 25:1, pp. 1-25.

Portmann, J. *When Bad Things Happen to Other People*. New York: Routledge, 2000.

Rizzolatti, Giacomo, L. Fadiga, L. Fogassi, and V. Ballege. "Resonance Behaviours and Mirror Neurons." *Archives Italiennes De Biologie*, 1999, 137, pp. 88-99.

Rizzolatti, Giacomo and L. Craighero. "Mirror Neuron: A Neurological Approach to Empathy." A working paper, University of Parma, Parma, Italy, 2004a.

_____. "The Mirror-Neuron System." *Annual Review of Neuroscience*, 2004b, 27, pp. 169-192

Rozin, Paul, Jonathan Haidt, and Clark R. McCauley. "Disgust." In Michael Lewis and Jeannette M. Haviland-Jones (eds.) *Handbook of Emotions*. New York: Guilford Press, 1993, pp. 575-594.

Rustichini. Aldo. "Neuroeconomics: Present and future." *Games and Economic Behavior*, 2005, 52:2, pp. 201-212.

Salovey, P. *The Psychology of Envy and Jealousy*. New York: Guilford, 1991.

Scheler, Max. *The Nature of Sympathy*; trans. by Peter Heath, intro. by W. Stark. New Haven, CT: Yale University Press, 1954.

Schoeck, H. *Envy: A Theory of Social Behaviour.* London: Secker & Warburg, 1969.

Sen, Amartya K. "Rational Fools: A Critique of the Behavioral Foundations of Economic Theory." *Philosophy & Public Affairs*, 1977, 6:4, pp. 317-344.

Silver, Maury and John Sabini. "The Perception of Envy." *Social Psychology*, June 1978, 41:2, pp. 105-117.

Smith, Adam. *The Theory of Moral Sentiments*, eds. D.D. Raphael and A.L Macfie. Oxford: Oxford

University Press, 1976.

_____. *The Correspondence of Adam Smith*, ed. E.C. Mossner and I.S. Moss. Oxford: Oxford University Press, 1977.

_____. *Lectures on Jurisprudence*, edited by R.L Meek, D.D. Raphael, and P.G. Stein. Oxford: Clarendon Press, 1978.

Smith, Richard H., Terence J. Turner, Ron Garonzik, Colin W. Leach, Vanessa Urch-Druskat, and Christine M. Weston. "Envy and *Schadenfreude*." *Personality and Social Psychology Bulletin*, 1996, 22, pp. 158-168.

Solomon, Robert C. "The Philosophy of Emotions." In Michael Lewis and Jeannette M. Haviland (eds.). *Handbook of Emotions*. New York: Guilford Press, 1993, pp. 3-15.

Stein, Edith (Saint Teresa Benedicta of the Cross Discalced Carmelite). *The Science of the Cross*, trans. by Josephine Koeppel. Washington, DC: ICS Publications, 2002.

_____. *On the Problem of Empathy*, 2nd ed. translated by Waltraut Stein, foreword by Erwin W. Straus. The Hague: M. Nijhoff, (1917) 1970.

Stigler, George J. and Gary S. Becker. "*DE Gustibus Non Est Disputandum.*" *American Economic Review*, March 1977, 67:1, pp. 76-90.

Sugden, Robert. "Beyond Sympathy and Empathy: Adam Smith's Concept of Fellow-Feeling." *Economics and Philosophy*, 2002, 18:1, pp. 63-87.

Sunstein, Cass R. and Richard Thaler. "Libertarian Paternalism is Not an Oxymoron." *University of Chicago Law Review*, Fall 2003, 70:4.

Sunstein, Cass R. and Martha C. Nussbaum (eds.). *Animal Rights: Current Debates and New Directions*. Oxford: Oxford University Press, 2004.

Thaler, Richard H. and Hersh M. Shefrin. "An Economic Theory of Self-Control." *Journal of Political Economy*, April 1981, 89:2, pp. 392-406.

Tomasello, Michael. *The Cultural Origins of Human Cognition*. Cambridge, MA: Harvard University Press, 1999.

_____, Malinda Carpenter, Josep Call, Tanya Behne, and Henrike Moll. "Understanding and Sharing Intentions: The Origins of Cultural Cognition." *Behavioral and Brain Sciences*, 2005, 28:5, pp. 675-691.

Tversky, Amos and Daniel Kahneman. "The Framing of Decisions and the Psychology of Choice." *Science*, 30 January 1981, 211, pp. 453-458.

Veblen, Thorstein. *The Theory of the Leisure Class*. New York: Modern Library, 1934.

Wilson, Edward O. *Sociobiology: The New Synthesis*. Cambridge, MA: Harvard University Press, 1975.